Three essays on financial analysts

Yaohua Li

Submitted in fulfilment of the requirements for

the Degree of Doctor of Philosophy

Adam Smith Business School

College of Social Sciences

University of Glasgow

March 2025

Abstract

This thesis comprises three chapters in related to the financial analysts. The first chapter focuses on the textual information in analyst reports. The second chapter examines the impact of analyst coverage. The third chapter investigates analysts' forecasting behavior in relation to the cultural characteristics of covered companies.

The first study in Chapter one examines the determinants and the impacts of analysts' tendency to provide similar textual information in their reports. To implement the tests, I collect a large sample of analyst report transcripts for the S&P 500 companies from 2015 to 2020. I introduce four factors to investigate the possible determinants of analyst report similarity, which include analysts' herding behavior, analysts' lack of ability, analysts' learning behavior, and the significant firm-related news. The regression results indicate that these four factors are positively associated with the analyst report similarity, which suggests that the four factors are likely to be the determinants of analyst report similarity. Next, I examine whether this textual similarity among analyst reports has implications for the market. The results indicate a negative association between the analyst report similarity and the short-term investor reactions, suggesting that market investors consider an analyst report as less informative if the report contains the textual information that is more similar to that in other prior analysts' reports. In the additional analysis, I find that this similarity likely hinders investors' understanding of analysts' quantitative outputs, including the earnings forecasts, stock recommendations, and price target forecasts. In addition, I find that the observed negative investor reaction to the analyst report similarity is stronger when firm managers have more incentives to withhold relevant information, but it is alleviated if the analysts are from larger brokers.

The second study in Chapter two investigates the influence of analysts on corporate governance in the context of corporate culture. I first examine the association between the analyst coverage of a firm and the score of the firm's culture. The baseline results show that the firm with higher level of analyst coverage is associated with a lower score of corporate

culture. This is consistent with the argument that analysts can impose short-term pressure on firms, resulting in a weak corporate culture. In further tests, I find that this negative association is stronger for the long-term oriented cultural values than other cultural values. Furthermore, to deal with the potential endogeneity problems, I first employ the two-stage least squares model based on a commonly used instrumental variable in this field, the expected coverage. The results suggest that the analyst coverage has a negative impact on the corporate culture. Moreover, I implement a quasi-natural experiment based on two exogenous shocks to analyst coverage, the brokerage closures and mergers. Consistent with the above findings, the results of the Difference-in-Difference model indicate that the analyst coverage could negatively influence the covered firm's culture. Taken together, these results are mostly consistent with the pressure effect that analysts impose short-term pressure on firms, increasing management myopia and resulting in a weak corporate culture. In additional tests, I find that analysts' negative effect on the corporate culture is alleviated when firms are covered by more experienced analysts, when firms are more likely to reach analysts' earnings forecasts, and when firms tend to have better corporate governance as captured by a higher competitive market.

The third study in Chapter three examines whether analysts' forecasting behavior is affected by the covered firm's information environment that is characterized by a strong integrity culture. I first investigate the association between analyst forecast boldness and the covered firms' scores of integrity culture. The baseline results indicate a positive relationship between the two variables, suggesting that analysts tend to issue bolder earnings forecasts when covering the firm with a stronger integrity culture. In further analysis, the results show that analysts' bolder forecasts for firms with a strong integrity culture are associated with decreased market reactions, indicating that analysts' bold opinions for these firms are regarded as less informative by market investors. Moreover, I find that the firm with a higher score of integrity culture is negatively associated with the number of analysts following the firm, suggesting that the analyst's service is less demanded for the firm with a stronger integrity culture. Additionally, other results show that analysts tend to issue less accurate earnings forecasts for firms with higher scores of integrity culture. Furthermore, I mitigate the potential endogeneity issue by conducting the two-stage least squares regression based on two instrumental variables: the average score of corporate integrity culture in an industry and the CEO age of the firm. Consistently, the regression results support the positive association between the firm's integrity culture and analyst forecast boldness. Finally, I introduce an alternative measure of the score of corporate integrity culture based on the analyst report transcripts. Consistent with the main findings, the regression results show that the corporate integrity culture is positively associated with analysts' forecast boldness. However, the association is not statistically significant, possibly due to the limited sample of analyst report transcripts.

Contents

List of tables

Acknowledgements

First and foremost, I want to express my immense and sincere gratitude to my supervisors, Professor Mark Aleksanyan, Professor Charalampos Stasinakis, and Professor Yannis Tsalavoutas, for their continuous support, encouragement, guidance, and assistance throughout the entire research process. This study could not have been completed without their help. It was an honor to work with them and study under their supervision.

Additionally, I want to thank all of the administrative staff at the Adam Smith Business School, who provided great assistance with my work and my life. I also want to extend my thanks to other staff members who provided helpful suggestions and insights for my research. Furthermore, I want to thank my friends who have supported me in both my studies and my personal life.

Last but not least, I am deeply grateful to my parents for their unconditional love, care, and support.

Author’s declaration

“I declare that, except where explicit reference is made to the contribution of others, that this dissertation is the result of my own work and has not been submitted for any other degree at the University of Glasgow or any other institution.”

Printed Name: Yaohua Li

Signature:

Introduction

This thesis is composed of three chapters that are related to the financial analysts. The first chapter concentrates on the textual information in analyst reports. The second chapter investigates the impact of analyst coverage on the covered firms' corporate culture. The third chapter examines analysts' forecasting behavior in related to the covered firm's information environment that is characterized by a strong integrity culture.

The financial analysts, as important information intermediaries, are responsible for collecting and analyzing the information about the covered firms, then providing their analysis results for other market participants. Analysts normally provide several analysis outcomes. On the one hand, they could generate various quantitative outputs. For example, analysts could make their predictions about the firm's future earnings, based on their analysis. They may also provide their recommendations about whether to buy or sell the stock of a company. The extensive studies have concentrated on such quantitative information and provided massive empirical findings about whether and how these analyst outputs could provide value for the market. On the other hand, analysts' important analysis outputs include the textual information in the analyst reports, which contain the detailed analysis of the covered firms. However, in compared with the numerous studies of analysts' quantitative outputs, the research into the textual information in analyst reports seems to be relatively scarce until recently. By introducing the textual analysis models, some recent studies provide much insight into the textual content in analyst reports. For instance, Huang et al. (2014) have conducted the machine learning approach to extract the textual opinions from a large sample of analyst report transcripts. They find that these textual opinions in analyst reports could provide incremental information for market investors beyond that in the quantitative outputs, such as the earnings forecasts and stock recommendations. Moreover, De Franco et al. (2015) have focused on the readability of analyst reports. Their results suggest that the lower readability could reduce the informativeness of analyst reports.

To provide new insight into the study of the textual information in analyst reports, I concentrate on another characteristic that is related to the textual information in analyst reports. Specifically, I study analysts' tendency to provide similar textual information in their reports. In related to this topic, prior studies observe that analysts, due to various reasons, tend to generate similar quantitative outputs (e.g., issuing similar earnings forecasts or stock recommendations) and such similarity could significantly influence analysts' value for the market (Clement and Tse, 2005; Jegadeesh and Kim, 2010). However, to the best of my knowledge, it remains an open question whether the similarity among analyst reports could be explained by relevant factors and whether the analyst report similarity has implications for the market.

In Chapter one, I investigate the determinants and effects of the analyst report similarity. To conduct these tests, I collect a large sample of analyst report transcripts for S&P 500 companies from 2015 to 2020. First, according to the prior studies, analysts' tendency to provide the reports with similar textual content is likely to be explained by the following four factors: analysts' herding behavior, analysts' poor ability to collect and provide new information, analysts' learning behavior, and the significant firm-related news. Consistent with this argument, the regression results show that the proxies for these four factors are positively associated with the analyst report similarity, suggesting that these four factors are likely to be the determinants of the similarity among analyst reports. Furthermore, I study the impact of the analyst report similarity by examining whether this textual similarity could influence the informativeness of analyst reports. The results show that the analyst report similarity is negatively associated with the short-term investor reactions immediately after the report date, indicating that market investors regard the analyst report as less informative if the report contains the textual information that is more similar to that in other analysts' prior reports. In the additional analysis, my results show that the market investors likely use the textual information in analyst reports to help understand the analysts' quantitative outputs (e.g., earnings forecasts). These quantitative outputs become less informative without new textual information in analyst reports to support them. Moreover, I find that the negative association between the analyst report similarity and the investor reactions is stronger when firm managers have more incentives to conceal relevant corporate information. In contrast, this negative relationship is found to be alleviated if analysts are from larger brokers. Finally, as a robustness test, I find that the analyst report similarity is negatively associated with the short-term abnormal trading volume.

My second chapter shifts the focus from the analyst reports to how financial analysts may affect the covered firms' corporate governance in the context of corporate culture. The corporate culture, which refers to the shared values and norms within the firm, has great importance in the corporate governance. According to the prior studies, the strong culture within the firm could improve employees' commitment, productivity, and self-discipline (Graham et al., 2022a). A strong corporate culture is also found to be associated with less short-termism among corporate executives (Quinn, 2018), higher stability during the financial crisis (Fang et al., 2023), and better performance during the COVID-19 (Li et al., 2021a). However, according to Graham et al. (2022a), they find from their survey that only a small percentage of the responding corporate executives believe that the culture in their firms is where it should be, despite the potential advantage of a strong corporate culture. And compared to the literature about the impact of corporate culture, the studies of the determinants of corporate culture is relatively deficient. My research in this chapter contributes to this field of study by investigating the role of an external agency, financial analysts, in the construction of a strong corporate culture. Based on the literature, the analysts likely have two opposite influences on the corporate culture. First, analysts could reduce the covered firm's information asymmetry and serve as external monitors, which may lead to a stronger corporate culture within the firm. In contrast, analysts are likely to impose short-term pressure on the firm (e.g., through their earnings forecasts), which may press the firm to generate instant profits to meet the earnings targets, leading to underinvestment in corporate culture.

In Chapter two, I investigate the association between the analyst coverage and the covered firms' corporate culture. To capture the corporate culture, I use the score of corporate culture developed by Li et al. (2021b) who conduct the textual analysis model on a large sample of public firms' conference call transcripts to obtain the score of corporate culture. I first examine the relationship between the analyst coverage and the score of the firm's culture. The baseline results show that firms with higher level of analyst coverage tend to have a lower score of corporate culture, which is consistent with analysts' pressure effect. Further results indicate that this possible negative influence is more concentrated on the long-term oriented cultural values (e.g., the culture of innovation) which are more aligned with the long-term corporate benefits rather than short-term profits. To deal with the potential endogeneity issues in baseline results, I first employ the two-stage least squares model based on an instrumental variable, expected coverage, which is proposed by (Yu, 2008). The results of this model suggest that the analyst coverage has a negative influence on the corporate culture. Furthermore, I implement a quasi-natural experiment based on two exogenous shocks, the brokerage closures and mergers, which are more likely to exogenously change the level of analyst coverage for a firm (Hong and Kacperczyk, 2010; Kelly and Ljungqvist, 2012). The results of the Difference-in-Difference model also indicate that the analyst coverage negatively affects the covered firms' corporate culture. Hence, the results from these two endogeneity tests are consistent with the baseline results and suggest that the observed negative association is likely causal. In additional analysis, I examine whether this negative association could vary with several analyst and firm characteristics. I find that analysts' negative impact on the corporate culture is alleviate for firms that are covered by more experienced analysts, for firms that are more likely to beat analysts' earnings forecasts, and for firms with a better corporate governance as captured by a higher competitive market.

My third chapter investigates whether analysts' forecasting behavior is affected by the covered firm's information environment that is characterized by a strong integrity culture. The financial analysts, as important information intermediaries, are required to collect and analyze the firm-related information, and to provide the analysis about the firm (e.g., earnings forecasts) for market investors. Therefore, financial analysts' behavior might be greatly affected by the corporate information environment. In this context, many prior studies have focused on analysts' forecasting behavior and provided empirical findings about the impact of corporate information environment. For instance, the prior literature finds that analysts tend to have less accurate earnings forecasts for firms with less readable annual reports (Lehavy et al., 2011) and for firms that restate their financial statements (Ye and Yu, 2017). However, it is not known from the literature whether the analysts' forecasting behavior is influenced by the corporate information that is characterized by a strong integrity culture. According to the prior studies, the corporate integrity culture represents a set of values and norms that emphasize integrity-related attributes (e.g., integrity, trust, honesty, transparency, and ethics). A strong integrity culture thus suggests that these integrity-related values and norms are widely shared and strongly held within the firm members, which could lead to a corporate information environment that is characterized by the above integrity attributes.

To fill the gap in this field, my research examines whether the firm's information environment that is characterized by a strong integrity culture could influence analysts' forecasting behavior. I concentrate on analysts' earnings forecasts, which, as discussed above, is one important output from financial analysts. Rather than the forecast accuracy, this research investigates the forecast boldness, which is also found to affect the value in analyst earnings forecasts (Gleason and Lee, 2003; Clement and Tse, 2005). According to the literature, a firm's integrity culture may have two opposite impacts on the analyst's forecast boldness. First, the firms with a stronger integrity culture likely reduce the information gap among analysts. Hence, analysts tend to issue forecasts that are less deviated from others. In addition, analysts might be more inclined to cover these firms. This could lead to higher analyst coverage and greater competition among analysts, which might encourage analysts to take herding behavior. In these two cases, analysts tend to issue less bold earnings forecasts. On the contrary, analysts may have bolder forecasts for firms with a stronger integrity culture. Specifically, the corporate information environment with higher integrity could make the analysts more certain about their own information and thus reduce their incentives to take herding behavior. In addition, the demand for analysts' service might be decreased due to the possibly lower information asymmetry between firms and investors. In this case, analysts might make more efforts to discover new information that is not easily available for investors. Analysts might also take the anti-herding behavior to differentiate themselves from others by strategically issuing more deviated forecasts, for the purpose of signaling their ability, impressing investors and generating publicity to attract clients. In these three cases, analysts tend to issue bolder earnings forecasts.

In Chapter three, I investigate whether analysts' forecasting behavior is influenced by the covered firm's information environment that is characterized by a strong integrity culture. I first examine whether the firms' integrity culture is associated with bolder earnings forecasts from analysts covering the firms. The results show that analysts tend to issue bolder earnings forecasts for the firms with higher scores of corporate integrity culture. To better understand the underlying explanation for analysts' forecast boldness, several further tests are conducted. Specifically, I find that analysts' bolder forecasts for firms with higher scores of integrity culture are regarded by the market as less informative, suggesting that this boldness is more likely to reflect analysts' anti-herding behavior. I also find that the firm's score of corporate integrity culture is negatively associated with the number of analysts following the firm, indicating that analyst's service is less required by investors for firms with a strong integrity culture. Moreover, the results show that analysts' forecasts for these firms tend to have lower accuracy. Taken together, the results so far are more consistent with analysts' anti-herding behavior of strategically differentiating themselves from others. Furthermore, to mitigate potential endogeneity concerns, I implement the two-stage least squares regression based on two instrumental variables: the average score of corporate integrity culture in an industry and the CEO age of the firm. Consistent with the baseline results, I find that analysts tend to have bolder forecasts for firms with a stronger integrity culture. Finally, I construct an alternative measure of the score of corporate integrity culture based on the analyst report transcripts. Consistently, the results show positive association between corporate integrity culture and analysts' forecast boldness, but this association is not statistically significant probably because of inadequate variation in the alternative measure across firms and years.

Collectively, this thesis makes several key contributions to the literature about financial analysts' practices and behaviors, which provides new insights into their roles as information intermediaries, the impact of their activities on corporate governance, and the factors that may influence their forecasting behavior.

The research in Chapter one shifts the focus from largely studied quantitative outputs (e.g., earnings forecasts and stock recommendations) to the textual content of analyst reports, offering new insights into analysts' tendency to exhibit similar behavior. By analyzing a large sample of analyst report transcripts, the study demonstrates that analysts' tendency of acting similarly extends beyond quantitative metrics to textual communication. The findings also indicate that textual similarity among reports reduces their informativeness, which highlights the importance of originality and uniqueness in textual content. The study further reveals that market investors value analysts' ability to discover new information not only beyond what firms disclose but also beyond what other analysts report. These findings extend the understanding of analysts' information intermediary role, emphasizing the importance of textual analysis in analyzing their behavior and its consequences.

The study in Chapter two contributes to the debate on the role of financial analysts by examining their impact on corporate culture. While analysts are mostly known to reduce information asymmetry and act as external monitors, this research reveals that their short-term pressure could lead to a weaker corporate culture. This study also complements prior studies on reduced innovation activities (e.g., fewer patents) by showing that analyst coverage could damage the underlying cultural mechanisms that encourage innovation. This study highlights the unintended negative consequences of analyst activities, demonstrating how external pressure from analysts can influence the construction of internal corporate culture.

The research in Chapter three explores corporate integrity culture as a fundamental aspect of firms' information environment and examines its influence on analysts' forecasting behavior. The findings show that analysts tend to issue bolder forecasts for firms with stronger integrity cultures, reflecting how the firm's information environment, characterized by a strong integrity culture, could influence the external stakeholders' perceptions and behaviors. Additionally, the study provides novel evidence of analysts' anti-herding behavior when covering such firms, as indicated by the less informative and less accurate forecasts, as well as the reduced demand for analysts' services. This implies that greater corporate transparency may reduce the perceived value of analysts' services, prompting them to take more aggressive strategies to maintain relevance.

1. Analyst Report Similarity

1.1. Abstract

In this research, I investigate the determinants and the impacts of analysts' tendency to provide similar textual information in their reports. To conduct these tests, I collect a large sample of analyst report transcripts for S&P 500 companies from 2015 to 2020. I first investigate the possible determinants of analyst report similarity. The regression results show that analysts' herding behavior, analysts' lack of ability, analysts' learning behavior, and the significant firm-related news are positively associated with analysts' tendency to discuss similar textual information in their reports, suggesting that these four factors are likely to be the determinants of analyst report similarity. Furthermore, I examine whether this textual similarity among analyst reports may influence the informativeness of the reports. The results indicate a negative association between analyst report similarity and the investor reactions, suggesting that market investors consider analyst reports as less informative when the reports provide textual information that is more similar to that in other analysts' prior reports. In additional analysis, I find that this similarity is detrimental to investors' understanding of analysts' earnings forecasts, stock recommendations, and price target forecasts. Moreover, this negative investor reaction is more pronounced when firm managers have more incentives to withhold relevant information, but it is moderated if analysts are from larger brokers. Overall, this study provides the large sample empirical findings about the determinants and influences of analysts' tendency to provide similar analyst reports.

1.2. Introduction

As information intermediaries, financial analysts are responsible for collecting and analyzing information about firms, then providing their analysis outcomes for market investors. An analyst typically produces several quantitative measures, such as the earnings forecasts, stock recommendations, and target price forecasts. Their important outputs also include the textual information in the analyst reports, which contains the detailed analysis of the covered firms. Compared to the relatively large amount of research that focus on analysts' quantitative outputs, especially upon the earnings forecasts and stock recommendations, relatively fewer studies have discussed the textual information in analyst reports.

With the development of computational technology and textual analysis tools, recent literature pays more attention to the textual information in analyst reports to conduct large sample research. For example, Huang et al. (2014) implement textual analysis tools upon nearly half a million analyst report transcripts for S&P 500 companies. They find that analyst reports bring market investors with more information beyond that in the earnings forecasts, stock recommendations, or target price forecasts. They also find that the textual information could help investors interpret these quantitative signals. In addition, De Franco et al. (2015) concentrate on one important textual attribute of analyst reports, that is, the readability of analyst reports and its effect on the informativeness of reports. They collect more than 300,000 analyst reports for public U.S. firms and find that more readable analyst reports increase the informativeness of these reports.

Although the recent studies provide valuable insight into the textual information in analyst reports, many textual characteristics in related to the analyst reports are not fully investigated. In this research, I focus on one of the textual attributes, the textual similarity, to investigate the determinants and impacts of analysts' tendency to provide similar textual information in their reports. To the best of my knowledge, this topic has not been studied previously in the research field related to the textual content in analyst reports. In contrast, prior studies on the analysts' quantitative outputs have conducted relatively more research into analysts' tendency of taking similar actions (i.e., issuing similar earnings forecasts or stock recommendations) (Clement and Tse, 2005; Jegadeesh and Kim, 2010).

Based on the prior literature, analysts' tendency of providing reports with similar textual content can be explained by at least four factors. First, analysts might have herding behavior, which refers to the tendency of an analyst to follow others and take similar actions, regardless of the analyst's own information (Scharfstein and Stein, 1990; Trueman, 1994; Hong et al., 2000). Prior literature suggests that individuals may follow others (herding) to share the blame because they want to keep their reputation and career, or they are not confident of their self-assessed ability (Hirshleifer and Teoh, 2003; Kadous et al., 2009; Jegadeesh and Kim, 2010). For example, the weak analysts might mimic the strong analyst's actions regardless of their own information, which leads to herding. Thus, if the analysts are taking herding behavior by following others, they are more likely to disregard their own private information and tend to provide similar textual information in their reports. In this case of analyst herding, an analyst report is less informative if it contains the textual information that is more similar to the prior analyst reports.

Second, even if analysts have no desire to take herding behavior, their ability to serve their information intermediary roles can vary. If an analyst is less capable of collecting and providing new information about firms, this analyst's report is more likely to contain the textual information that is stale to the market and is more similar to the textual content in prior analysts' reports. Since market investors highly value analysts' efforts to discover new information (Ivković and Jegadeesh, 2004; Livnat and Zhang, 2012; Huang et al., 2018), higher similarity in the textual information in analyst reports due to the analysts' poor ability can reduce the informativeness of these reports.

Third, analysts might be able to learn from the actions of prior analysts. For example, an analyst can analyze the useful information contained in prior analysts' earnings forecasts to obtain the analyst's own forecasts, which is found to be associated with higher accuracy of earnings forecast (Clement et al., 2011; Kumar, 2022). With respect to the textual information in analyst reports, an analyst might extract useful information in prior analysts' reports to complement his own information set. Consequently, the textual information in the analyst report could be more similar to that in reports from other prior analysts. In this case, the analyst report similarity due to analysts' learning behavior reflects analysts' efforts to collect and analyze corporate-related information, which is not supposed to reduce the informativeness of analyst reports.

Finally, the analysts can take similar actions because of the significant firm-related news (Graham, 1999; Welch, 2000; Jegadeesh and Kim, 2010). In this context, prior literature finds that market investors think highly of analysts' role to interpret or confirm the information immediately after the release of a firm-related event (Livnat and Zhang, 2012; Huang et al., 2018). For example, after the earnings announcement by a firm, a great percentage of analysts covering the firm would discuss the information related to the earnings announcement. In this condition, analysts tend to provide similar textual information in their reports because of their interpretation, confirmation, and analysis of the common public information from the earnings announcement. Hence, the similar textual information in analyst reports might serve to interpret, justify, or confirm the prior information, which might not decrease the informativeness of these analyst reports. On the contrary, if the market regards this as simply repetition of prior information that provides very little new information, the market is likely to consider these reports as less informative.

To examine the determinants and consequences of analyst report similarity, I collect a large sample of 571,661 analyst report transcripts for S&P 500 companies from 2015 to 2020, which are downloaded from the Investext through Thomson One database. To obtain the measure of textual similarity among analyst reports, I calculate the average value of cosine similarity scores between the word frequency vector of an analyst report and those from other prior analyst reports. Moreover, I calculate the investor reactions to capture how informative an analyst report is for the market investors. Intuitively, a more informative analyst report tends to stimulate higher investor reactions (Huang et al., 2014). Specifically, I calculate the absolute value of abnormal stock return in the three-day period after the release of each analyst report.

Based on the regression results, I find that analysts are more likely to discuss similar textual information in their reports when analysts are likely to be the follower analysts and when analysts issue earnings forecasts that deviate less from the most recent prior consensus forecasts, which is consistent with that the analyst report similarity is positively associated with analysts' herding behavior. In addition, the results indicate higher textual similarity among analyst reports when analysts provide less new information compared to their own prior reports, and when analysts have lower frequency of revising their earnings forecasts in a year. This is consistent with the conjecture that if the analyst lacks the ability to collect and provide new information, this analyst is more likely to provide the textual content in the report that is more similar to that in other analysts' reports. Moreover, the results show that the textual content in an analyst report is more similar to that in other analysts' prior reports when these prior analyst reports likely provide more information, which indicates analysts' learning behavior. Additionally, I find higher similarity among analyst reports when greater number of reports are clustered around the report of interest, and when the absolute value of the prior ten-day abnormal stock return is higher. This suggests that the significant firm-related news could explain analysts' tendency to provide similar textual information in analyst reports.

I next examine the impact of the analyst report similarity on the informativeness of the reports as captured by the short-term investor reactions. It is observed from the results that the analyst report similarity is associated with lower value of the three-day absolute abnormal return, after controlling for a set of possible determinants of the short-term market reactions. As suggested by the prior study (De Franco et al., 2015), this short-window event research method helps construct a causal relationship. Thus, the results are mostly consistent with that the analyst report similarity could reduce the informativeness of analyst reports.

The results from additional analysis show that when analysts tend to provide similar textual information in their reports, these analysts' quantitative outputs (i.e., the earnings forecasts, stock recommendations, and price target forecasts) become less informative for investors. This could suggest that analysts' textual information in reports is valuable for investors to understand the analysts' quantitative outputs, and that these quantitative figures are less informative without new textual information in analyst reports to support them.

Moreover, the observed negative association between the analyst report similarity and the investor reactions is more pronounced when the firm's actual earnings announced around the report date is lower than the analysts' expectation, suggesting that investors might place more reliance on analysts' efforts to provide relevant new information when firm managers have more incentives to hide corporate information. In addition, it is found that the investors' negative reaction to the analyst report similarity is moderated if analysts are from larger brokers, indicating that these analysts provide more informative content in their reports. As a robustness test, I construct an alternative measure based on the abnormal trading volume to capture the investor reaction. The results based on this new measure are consistent with the baseline results.

This study makes several contributions to the extant literature. First, this study adds to the broad literature about analysts' tendency to exhibit similar behavior by shifting the focus from quantitative outputs to the textual content of analyst reports. Prior literature has mostly studied the factors that lead to similarity in analysts' quantitative outputs, particularly analysts' tendency to issue similar earnings forecasts or stock recommendations (Clement and Tse, 2005; Bernhardt et al., 2006; Jegadeesh and Kim, 2010; Clement et al., 2011; Hahn and Song, 2013; Cao et al., 2023). According to these previous studies, analysts are likely to provide similar quantitative outputs when they are trying to follow others (i.e., herding behavior), when they lack the ability to discover new information, when they are trying to learn from others, and when there is important firm-related news around the release of the analyst report. Compared to the previous studies, my research provides new insight into this study field by concentrating on the similarity among textual information in analyst reports instead of the quantitative outputs. Based on a large sample of analyst report transcripts, this study investigates whether analysts tend to provide similar textual information in analyst reports due to the above reasons, which sheds light on the underlying mechanisms driving this behavior. Furthermore, this research examines how the analyst report similarity might affect the informativeness of analyst reports, offering new insights into the consequences of such convergence. These findings not only broaden our understanding of analysts' behavior, but also emphasize the importance of textual analysis in analyzing analysts' behaviors that quantitative metrics alone might not reveal.

Second, this study enriches the growing field of research in related to the textual information in analyst reports (Twedt and Rees, 2012; Huang et al., 2014; De Franco et al., 2015; Huang et al., 2018). In particular, as suggested by prior literature (Asquith et al., 2005; Huang et al., 2014; Imam and Spence, 2016), the textual content in analyst reports provides important information for investors even beyond the quantitative outputs such as earnings forecasts and recommendations. Correspondingly, my research advances the previous literature by concentrating on a critical yet underexplored attribute of analyst reports, that is, the textual similarity among these reports. By computing and analyzing textual similarity across a large sample of analyst reports, I provide empirical evidence on the determinants of such similarity and its impact on the informativeness of analyst reports. The findings suggest that textual similarity reduces the informativeness of reports, highlighting the importance of originality and uniqueness in the textual content of analyst reports. This research provides new insights into how analysts' textual communication influences investor decision-making. The contribution is more pronounced given the relatively limited attention paid to textual similarity compared to the extensive research on quantitative outputs.

Finally, this research extends the study of analysts' roles as information intermediary. Prior literature suggests that analysts provide value by discovering private information or interpreting public information (Ivković and Jegadeesh, 2004; Chen et al., 2010; Livnat and Zhang, 2012). These studies argue that analyst forecast revisions following (preceding) earnings announcements mostly reflect their information interpretation (discovery) role, and that both roles are valued by market investors. More recently, Huang et al. (2018) extend this literature by using the textual analysis method upon analyst report transcripts to explicitly capture analysts' discovery of new information. They show that in the two-day period after firms' earnings calls, investors value analysts' discovering of new information that is not disclosed in firms' documents (i.e., earnings calls transcripts). My research adds to the above literature by examining the role of textual content in conveying new information. Specifically, I show that an analyst report becomes less informative if its textual content is more similar to that in prior analyst reports, which is consistent with the view that analysts' role of discovering new information is important for investors. Furthermore, this study extends the literature by showing that the market investors value analysts' ability to discover new information, not only beyond what firms disclose, but also beyond what other analysts report. In addition, unlike Huang et al. (2018), who focus on the two-day period after earnings calls, my research examines analyst reports throughout the year, offering a more comprehensive understanding of analysts' information intermediary role over time.

1.3. Literature review and hypotheses development

1.3.1. Textual information in analyst reports

A great number of prior studies in the area of financial analysts has focused on the quantitative outputs of analysts, particularly the earnings forecasts and stock recommendations, and examined whether these quantitative outputs provide value to the market. Comparatively, relatively fewer studies have implemented large sample tests to investigate the value of textual information in analysts' reports.

More recently, with the development of the computation technology and textual analysis tools, the study of the textual content of a large sample of analyst reports is enabled. Recent studies show that the textual content in analyst reports provides significant information to investors, beyond the quantitative outputs, such as earnings forecasts and recommendations. Asquith et al. (2005) collect nearly 1,000 analyst report transcripts and extract the textual information in each report transcript that is related to analysts' justification of their stock recommendations. They manually read these materials to decide whether the remarks in these texts are related to any of the 14 given categories (e.g., revenue growth, earnings growth, and cost efficiencies). Then they construct a variable called "strength-of-arguments," calculated as aggregating the number of positive remarks less the number of negative remarks that are related to the 14 categories in each report. Their results show that, after controlling for earnings forecasts, recommendations, and target prices, the strength-of-arguments of analyst reports is significant in explaining five-day abnormal returns surrounding the report date. These results suggest that the textual information in analyst reports does provide value to market investors. Consistent with this finding, Twedt and Rees (2012) focus on nearly 2,000 initiation reports by analysts and calculate the tone and complexity of these reports. They find that these two textual attributes of the initiation reports have significant impact on the short-term stock return, after controlling for the effect of earnings forecasts and stock recommendations. This suggests that the report tone and complexity contain significant incremental information to the market investors compared to the quantitative outputs such as earnings forecasts and stock recommendations.

In addition, Huang et al. (2014) explore a much larger sample of 363,952 analyst reports issued for S&P 500 during 1995 to 2008. They use a textual analysis tool, the naive Bayes classification, to capture the textual opinions (i.e., the tone) of analyst reports. Based on the regression method, they find that the textual opinions in analyst reports have significant impact on the abnormal stock return, after controlling for important quantitative outputs (i.e., the earnings forecasts, stock recommendations, and target price forecasts) and other controls. Their results suggest that the textual discussions in analyst reports provide investors with information beyond that in the quantitative outputs. Their analysis further indicates that the textual content in reports helps investors interpret these quantitative outputs and that the textual information is more informative for investors if it contains negative news. In addition, De Franco et al. (2015) concentrate on one of the textual attributes, that is, the readability of analyst reports. They examine the effect of analyst report readability on the informativeness of these reports for market investors. Their results show that the more readable analyst reports are associated with higher abnormal trading volume for the three-day period around the analyst report. Their results suggest that more readable textual information in analyst reports could cause the investors to increase their trades.

The more recent study of Huang et al. (2018) employs a topic modeling methodology

(LDA) to extract topics from the textual content of a large sample of analyst reports that are release immediately after the quarterly earnings conference calls for S&P 500 companies during the period of 2003 to 2012. They split the textual information in analyst reports into two kinds of information (i.e., analysts' discovery of new information and interpretation of public information in conference calls) to see how market investors react to them differently. Their results show that the market values analysts' efforts to provide both kinds of information in their reports and puts a greater weight on analysts' interpretation of public information during the short-term period after earnings conference calls.

Based on the literature above, the textual information from reports not only justifies analysts' choices of their earnings forecasts and recommendations, but also equips investors with the value-relevant information about the firms, based upon which the analysts serve their role as information intermediary. However, compared to the great number of prior studies on earnings forecasts or stock recommendations, the study on the textual information in analyst reports is still under development. Furthermore, despite the several studies on incremental value of the analyst reports compared to the quantitative outputs, few have studied whether some of the textual attributes of the content in analyst reports might affect the informativeness of these reports.

1.3.2. Analyst information roles

In order to understand whether and how analysts can contribute to well-functioning capital markets, the prior literature has discussed different informational roles of analysts. It is noted from the literature that analysts could provide value and serve their information intermediary role at least by serving their information discovery role and information interpretation role: by collecting and discovering the private information that investors cannot easily access; and by interpreting and analyzing the public information that cannot be easily understood or processed.

Specifically, Ivković and Jegadeesh (2004) assume that analysts serve the two informational roles: information interpretation role and information discovery role. First, analysts serve their information interpretation roles by analyzing and clarifying existing public information, and by offering their own opinions. Meanwhile, analysts could collect and generate useful information that is not readily available to an investor, implying their information discovery roles. They classify analyst revisions of earnings forecasts and stock recommendations according to their timing relative to earnings announcements, with the key assumption that analyst revisions in the one-week period immediately after corporate earnings announcements reflect analysts' information interpretation role and that analyst revisions in other period indicate analysts' information discovery role. Consistent with such assumption, they find a sharply higher frequency of earnings forecast revisions and recommendation revisions around the earnings announcements date (EAD) than on other days. In particular, they observe 26.35% of all forecast revisions and 20% of all stock recommendation revisions on the EAD and the following two days. They find that the forecast revisions are less informative during the one-week after the earnings announcement, indicating that analysts' value are mainly from their information discovery role rather than interpretation role.

Chen et al. (2010) extend the prior research by assuming that analysts serve both information discovery role and interpretation role at the same time. They propose that if analysts primarily serve their information discovery roles, then analyst reports pre-empt the effect of subsequent earnings announcements. In contrast, the analysts' information interpretation role is likely to reinforce the impact of earnings announcements. Using equity market reaction to capture information content, they find that the information content in analyst forecasts issued prior to EAD is negatively associated with information content in earnings announcements, consistent with analyst research pre-empting the earnings report. Conversely, the information content in analyst forecasts issued immediately after EAD is positively associated with information content in earnings announcements, consistent with analyst interpreting the earnings report. In addition, the information interpretation role is more important when firm's operation is more complex. Their results support the arguments from Ivković and Jegadeesh (2004) that during the one-week period after earnings announcements, analysts' interpretation role (reinforce) dominates, while the discovery role (pre-empt) dominates in other period.

Livnat and Zhang (2012) take into consideration other important corporate information like 10-K, 10-Q, and 8-K to study these two roles around the corporate disclosures. They further include the prompt forecast revisions that are issued in the three-day period after these corporate disclosures, unlike the previous studies that exclude these prompt revisions from the sample. They presume that these prompt revisions after 8-K are more likely to reflect analysts' interpretation role and that the analyst revisions in other period tend to represent analysts' information discovery role. Their results show that analysts' forecast revisions issued immediately after corporate disclosures is associated with stronger market reactions than other revisions, which suggests that analysts' timely information interpretation of corporate public information is more informative for market investors.

A more recent research of Huang et al. (2018) use LDA model to separate the textual content in analyst reports into analysts' discovery of new information and their interpretation of public information. Specifically, they conduct the LDA model on transcripts of analyst reports and corporate conference calls to identify the topics that are discussed in each transcript. According to their method, the LDA topics that are discussed in both analyst reports and the immediately preceding corporate conference calls tend to reflect analysts' interpretation of public information, while the new topics in analyst reports are more likely to represent analysts' discovery of new information not disclosed in corporate conference calls. Their method explicitly measures the extent of these two information roles. Additionally, they measure the extent of analysts' simple confirmation of public information. Their results show that during the two-day period after conference calls, investor reactions to both roles are significant and investors place a greater value on the interpretation role. In addition, the investor reaction to information interpretation is more pronounced when processing cost of conference call information increases. Finally, they find that investors react to the content of analyst reports which simply confirms the information in conference calls. Their results suggest that during the period right after significant public information, investors value all three information roles and place a higher weight on analysts' interpretation roles.

Overall, most of prior literature employ analysts' forecast revisions and the subsequent market reactions to study analysts' information roles, rather than explicitly measure analysts' discovery of new information or interpretation of public information. Huang et al. (2018) extend it by conducting the textual analysis upon a large sample of analyst report transcripts. They show that analysts' information interpretation roles dominate in the two-day period after conference calls. However, it remains an open question whether this could apply to the period other than the two days after conference calls. Furthermore, these previous studies mostly investigate analysts' information discovery role by concentrating on whether analysts could discover private information that is not disclosed by companies. However, the textual content in analyst reports also provide valuable corporate information for market investors. What is not examined in the prior research is whether market investors value analysts' efforts to collect new information that is not provided in other prior analysts' reports.

1.3.3. Herding among analysts

The financial analysts are regarded as one of the most important information intermediaries in the stock market, who provides analysis about the covered firms for other market participants. However, analysts might not always provide objective and unbiased information. They may generate biased analysis outcomes because of analysts' herding behavior. Herding refers to the tendency of analysts, who make their own individual decisions, to follow other analysts' actions, even if such behavior cannot be fully justified by their own information (Scharfstein and Stein, 1990; Trueman, 1994; Jegadeesh and Kim, 2010). Due to this behavior, analysts might release similar quantitative outputs, such as the earnings forecasts and stock recommendations.

One important motivator to analyst herding behavior is the incentives. The herding behavior could be rational from the perspectives of analysts, regardless of whether such behavior could damage the interest of other market participants. The prior study on a similar agent, the firm manager, by Scharfstein and Stein (1990) suggests that firm managers may follow other managers' decision-making process, for the purpose of protecting their own reputation by sharing the blame with others. Therefore, if an investment project turns out to be unprofitable, this should not have a big impact on the manager's reputation because others make the same mistakes. Another research by Stickel (1990) concentrates on financial analysts and provides empirical evidence that analysts' career concern could influence their herding behavior. In particular, Stickel (1990) finds that the forecast revision of an analyst is more likely to be affected by the changes in prior consensus forecasts if this analyst is not a member of the *Institutional Investor*'s "All-American Research Team." Stickel (1990) further argues that herding (low-innovation) forecasts contain less information than bold (high-innovation) forecasts, because the herding forecasts are more likely to reflect the information revealed by the prior revisions of other analysts.

Consistently, Trueman (1994)'s model indicates that analysts with worse forecasting ability are more likely to exhibit herding behavior by issuing forecasts that are closer to other analysts, because analysts' forecasting ability is among the most important determinants of their compensations. In addition, Hong et al. (2000)'s results suggest that inexperienced analysts have bigger career concern, and thus are more likely to take herding behavior. In particular, they find that inexperienced analysts are more likely to be terminated for inaccurate forecasts than more experienced analysts, and that inexperienced analysts tend to herd by deviating less from the consensus forecasts.

Clement and Tse (2005) provide further evidence that both analysts' career concern and their self-assessed ability could influence analyst herding behavior. They show that the likelihood of issuing bolder forecasts is greater for (1) analysts with higher prior accuracy, (2) analysts from bigger brokers, (3) analysts that have more frequent forecast revisions, and (4) analysts with more general experience. On the contrary, the likelihood of herding behavior among analysts becomes higher for analysts who follow a greater number of industries, possibly because analysts are less capable of making analysis and forecasts when they cover too many industries. Moreover, they find that analysts who are from smaller brokers or cover a greater number of industries are not penalized (i.e., being terminated) for issuing bolder forecasts than others. Yet these analysts still have herding behavior, suggesting that analysts' self-assessed ability might have an additional impact on herding behavior compared to analysts' career concern.

Consistent with the prior findings, some more recent studies have found that the likelihood of herding behavior among analysts increases with analysts' career concern and their ability. For example, Jegadeesh and Kim (2010) find that analysts are more likely to herd when conveying negative corporate information (by downgrading their stock recommendations) to the market, since this might damage the relationship between analysts and firm managers. Nolte et al. (2014) observe greater herding behavior among analysts during the period of macro banking pessimism (as captured by the ratio of negative to positive news of firms in the banking sector) when the job security of analysts is lower due to the downsizing and cost cutting. Hirshleifer et al. (2019) and Jiao (2024) show that analysts are less likely to have bold forecasts when they are more fatigued (issuing too many forecasts in one day) and are thus, less capable of making forecasts. The results from He et al. (2020) indicate that analysts from the analyst teams with more clear hierarchy tend to outperform others by generating more accurate forecasts, which consequently reduces their incentives to herd.

Based on these findings, other recent studies concentrate on the effect of corporate information environment on analysts' herding behavior. Frijns and Huynh (2018) find increased herding behavior among analysts when the covered firms have negative media sentiment as captured by the negative tone of the media news, and when these firms have higher information uncertainty as measured by the disagreement in the media news. These results are also consistent with the prior findings that analysts could have herding behavior when they are unwilling to damage their relationships with firm managers and when they are less capable of making forecasts. Similarly, Leece and White (2017) show that analysts are more likely to have herding behavior for firms with higher percentage of ownership held by the transient institutional investors. Specifically, the incidence of the short-term institutional investors is positively correlated with the opaqueness of firms' information environment, and such opaqueness increase the difficulty for analysts to make forecasts, which leads to herding behavior. In addition, the results from Wen and Tikoo (2020) demonstrate that the tendency to herd among analysts is greater when the covered firms pursue more unique corporate strategies, because such uniqueness could impose higher information processing cost on analysts and make it more difficult for analysts to make forecasts.

Apart from the above analysts' incentives that are mostly related to analyst career concerns, some of the other determinants of analyst herding behavior involve analysts' personal attributes. Jiang et al. (2015) argue that analysts' preferences for Republican Party is associated with conservative traits. Their results show that the analysts preferring the Republican Party tend to have smaller revisions and less bold forecasts. Christoffersen and Stæhr (2019) and Cleary et al. (2020) find that less risk-tolerant analysts tend to have herding behavior by deviating less from the consensus. The results from Cao et al. (2023) exhibit that analysts are more likely to issue bold forecasts if they are from the countries with a stronger individualistic culture that encourage differentiation between individuals and believe individuals make a positive contribution to the society with their distinctive characteristics.

The above prior studies suggest that analysts might exhibit herding behavior by following others due to different reasons that are unrelated to the covered firms, which could be detrimental to analysts' role to provide unbiased analysis outputs (except the final case where analysts receive correlated information). Though intuitively, the prior literature provides further empirical evidence that analysts' herding behavior decrease the information provided to the market. For example, Clement and Tse (2003) and Gleason and Lee (2003) document smaller stock return reactions for herding forecast revisions than for bold forecast revisions. This is consistent with their argument that herding (low-innovation) forecasts tend to contain less information than bold forecasts. Clement and Tse (2005)'s research provides further explanation for these results. They find that herding forecasts are less accurate than bold forecasts. Thus, it is likely that analysts' herding forecasts provide less relevant information to investors than bold forecasts. Similarly, Jegadeesh and Kim (2010) focus on the analysts' stock recommendations and find that the stock reaction to analysts' recommendation revisions is stronger when the revised recommendations move away from the consensus than when they move toward the consensus recommendation.

In addition, a recent research of Da and Huang (2020) use Estimize.com to directly examine the effect of herding behavior on the information environment. Estimize.com is an open web-based platform founded in 2011 on which users can make earnings forecasts. By using blind experiments, they find that forecasters reduce the weight on their own private information after they observe other's forecasts. They find that this behavior of following other forecasters (i.e., herding behavior) improves the accuracy of an individual forecast, but it makes the consensus less accurate. After decomposing the consensus forecast errors into individual forecast errors and individual forecast diversity, their results show that despite the higher accuracy in individual forecasts, the reduced diversity makes the consensus forecasts less accurate. These results suggest that the herding behavior among forecasters reduces the accuracy of the consensus forecasts due to the less diverse opinions.

Overall, based on the discussion above, analysts might exhibit herding behavior by following others' actions due to different reasons (e.g., career concern and personal attributes) that are unrelated to the information about the covered firms. Under the circumstance, analysts' role to convey unbiased information is compromised by this tendency toward herding. Although prior literature suggests that analysts' herding behavior leads analysts to generate similar earnings forecasts and stock recommendations, what is not clear from the literature is whether such herding behavior is associated with higher textual similarity among analyst reports.

1.3.4. Other reasons of behavioral similarity

Other than the extensively studied herding behavior among financial analysts, analysts might take similar actions because of other reasons. First, according to the previously discussed literature about analysts' information roles, analysts could provide value for the market investors by discovering new information that is not easily accessible for investors (Ivković and Jegadeesh, 2004; Chen et al., 2010; Livnat and Zhang, 2012; Huang et al., 2018). Based on these prior findings, if an analyst lacks the access or ability to discover and collect valuable new information, he or she is likely to provide the stale information that is already reflected in other prior analysts' reports. Ceteris paribus, the informativeness of the content in the analyst report is lower for market investors since the analyst report contains less new information.

Second, the analyst report similarity might be affected by the following factor. Specifically, one can argue that an analyst may learn from the actions of preceding analysts, which could lead to the quantitative outputs more similar to those of other previous analysts. For example, Clement et al. (2011) investigate analysts' ability to learn from the actions of market investors and other analysts when making their own forecasts about corporate earnings. They focus on analysts' revisions of earnings forecasts immediately after the corporate earnings announcement and find that these revisions are significantly correlated with the market reaction to the earnings announcement and to the revisions of other analysts, especially when these two indicators tend to be more informative. Furthermore, their results show that these analysts who extract information from the two indicators tend to achieve higher forecast accuracy and trigger stronger market response. These results are mostly consistent with that analysts may learn from other analysts to improve their own forecasts. A more recent research by Kumar (2022) extends this by investigating analysts' social learning behavior that an analyst covering a firm could learn from other peer analysts following other firms that the analyst covers (i.e., other firms in the analyst's coverage portfolio excluding the focal firm). Consistent with the prior literature, Kumar (2022)'s results show that analysts revise their earnings forecasts by looking at the forecast errors of other peer analysts for other firms and that this learning from peers increases the analyst's forecast accuracy.

The above empirical findings are consistent with analysts' learning behavior that analysts could extract information from the actions of other analysts. For example, analysts may attend the firm's conference call and observe what are the concerns of other analysts. In this context, analysts might update their information set accordingly and generate their reports which contain the textual information that is more similar to that in other analysts' reports. Moreover, compared to the cases where an analyst tries to follow others (i.e., herding behavior) and where the analyst lacks the ability to discover and provide new information, the similarity due to the analyst's learning behavior might not be regarded as less informative by market investors since the analyst is trying to learn from others and extract information from their actions.

Finally, analysts might tend to act more similarly when they receive relatively similar information. Prior literature suggests that, under such circumstance, analysts tend to act similarly. For instance, Graham (1999) finds that analysts are more likely to have similar recommendations when the informative private signals are positively correlated across analysts. Such information correlation is captured by the scaled cross-sectional standard deviation of private forecasts of the three-month T-bill rate, given that analysts likely use the private forecasts of the T-bill rate as useful inputs to predict future market movement. Another research of Hahn and Song (2013) concentrates on the influence of the adoption of Regulation Fair Disclosure which prohibits companies from selectively disclosing material information to favored analysts. Consistent with the previous empirical evidence, they find that, after the RFD, analysts are more likely to release their earnings forecasts immediately after the firm's earnings announcements, and these earnings forecasts become more converging. This suggests that analysts depend more on the company's disclosed information from earnings announcements after the RFD and such common corporate information leads to similar earnings forecasts.

Although the arrival of company-related information could lead analysts to provide similar textual information in their reports, the possible influence of such similarity is not apparent. Based on the previous discussion about analysts' information roles, market investors may sometimes find analysts' interpretation or confirmation of corporate information to be incrementally informative, even beyond the discovery of new information. Hence, in the cases where analysts tend to provide similar textual information in reports after receiving significant corporate information, the market reaction of investors might be twofold. In particular, if the investors regard the interpretation or confirmation of the corporate information as more important, the investors might not consider the similar textual information in analyst reports as less informative. However, if the market investors tend to perceive this as less important than the discovery of new information, the market might regard these reports as less informative.

1.3.5. Hypotheses development

Based on the discussion above, the prior studies suggest that the textual content in analyst reports provides valuable corporate information to market investors, beyond the information contained in analysts' quantitative outputs (e.g., earnings forecasts, stock recommendations, and target price forecasts). In this study, I concentrate on one relevant textual attribute, the textual similarity, of the content among analyst reports.

In this section, I develop the hypotheses about whether and how the analyst report similarity could influence the informativeness of these reports. Rather intuitively, an analyst report is supposed to be less informative if the textual information in the report is more similar to the textual content in the prior analysts' reports. Four factors that possibly increase the textual similarity among analyst reports are discussed below.

According to the discussion above, analysts may have herding behavior by following other analysts' actions regardless of their own information, because of reasons that are not related to the covered firms' information. In particular, previous studies find that analysts are more likely to take herding behavior if they are more concerned about their career (Hong et al., 2000; Clement and Tse, 2005; Frijns and Huynh, 2018). As a result, such herding behavior among analysts leads them to generate similar quantitative outputs (e.g., earnings forecasts and stock recommendations). This herding behavior could have similar implications with respect to the qualitative outputs of analysts (i.e., the textual information in analyst reports). Specifically, if an analyst chooses to follow other prior analysts' actions and disregard their own private information, this analyst is more likely to provide information that is similar to prior analysts. Under this circumstance, an analyst report is less informative for market investors if the textual information in it is more similar to that in the prior analyst reports, because this report is more likely to contain the information that is already provided by prior reports instead of new information.

In addition, based on the discussion previously, the analysts could provide value by discovering new information that is not readily available for market investors (Ivković and Jegadeesh, 2004; Chen et al., 2010; Livnat and Zhang, 2012; Huang et al., 2018). Accordingly, if an analyst lacks the ability to collect valuable new information, he or she may provide stale information that is already disclosed in prior analyst reports. Ceteris paribus, if an analyst report contains less new information, the informativeness of the report is possibly lower for the market investors.

Furthermore, as mentioned in the prior section, an analyst might learn from the preceding analysts (Clement et al., 2011; Kumar, 2022). Specifically, based on these empirical findings, an analyst could extract useful information from the actions of other analysts. As a result, the textual content in the analyst's report might be more similar to that in other prior analysts' reports. In this context, the higher similarity might not be regarded by market investors as less informative, compared to the scenarios where an analyst is trying to simply follow others (i.e., herding behavior) and where the analyst lacks the ability to collect and provide new information.

Moreover, according to the literature in the above section, in the situations where analysts likely receive common information about the firm, they might seem to provide similar textual information in their reports (Graham, 1999; Hahn and Song, 2013). In this context, if the market investors regard such interpretation or confirmation of the corporate information as more informative than the discovery of new information, the investors might not perceive the textual information in these reports as less informative. In contrast, if the market investors regard such information roles as less important which provides very little information, they would consider these reports as less informative.

Overall, the above discussion could lead to the first hypothesis:

Hypothesis 1. An analyst's tendency to provide textual information that is more similar to prior analyst reports is positively correlated with the analyst's herding behavior, the lack of ability to collect and provide new information, the learning behavior, and the effect of important firm-related news.

Furthermore, with respect to the informativeness of analyst reports, the above discussion provides useful insight. If an analyst report tends to contain similar textual information with the prior reports due to the analyst's herding behavior of following prior analysts or the analyst's lack of ability to collect and provide new information, the analyst is less likely to provide valuable information in the report, which reduces the informativeness of this report. In addition, analysts may provide similar information in their reports when there is significant firm-related news. If the market regards this similarity as providing little new information, the informativeness of these reports for market investors could be low. I use the market investor response to capture the informativeness of the information in analyst reports. These intuitions lead to the following hypothesis:

Hypothesis 2a. An analyst report is less informative if the report contains the textual information that is more similar to that in reports by prior analysts.

In contrast, according to the above discussion, the analyst report similarity might be derived from analysts' learning behavior and the emergence of significant firm-related news. Specifically, an analyst may extract useful information from other prior analysts' actions. When there is significant firm-related news, analysts are likely to make efforts to interpret and analyze the information in related to the news, which could lead to relatively more similar textual information in analyst reports. In these cases, analysts are less likely to issue biased analysis outputs, unlike analysts' herding behavior or the lack of analysts' ability to collect and provide new information. Thus, the market investors might not consider the textual information in these reports as less informative, leading to the following hypothesis:

Hypothesis 2b. The informative of an analyst report is not associated with whether the report contains the textual information that is more similar to that in reports by prior analysts.

1.4. Research design

1.4.1. Sample selection

Several databases are used for the collection of the necessary data for the analysis in this study. The analyst forecast data is obtained from the Institutional Brokers' Estimate System Database (I/B/E/S). The accounting data of companies is collected from the COMPUSTAT database, while the stock market price data comes from the Center for Research in Security Prices Database (CRSP). I next obtain the word lists developed and used by prior research (Loughran and Mcdonald, 2011, 2015; Wang, 2020), for the purpose of measuring the tone and other characteristics of the textual content in analyst reports. The most recent version of this provided word list is updated in 2021 and it recognizes seven groups of words: Negative, Positive, Uncertainty, Litigious, Strong Modal, Weak Modal, and Constraining. This word list is adjusted for financial language and is more appropriate than other word lists (Bellstam et al., 2020).

As a starting point for collecting analyst report transcripts, I identify the sample of firms that were listed on the S&P 500 Index at any point during the period of 2015 to 2020. The sample of S&P 500 companies are frequently used in recent studies on analyst reports (Huang et al., 2014; Huang et al., 2018; Bellstam et al., 2020). Choosing these firms is likely to have several advantages. Specifically, the companies with high-profile are almost always covered by brokers and are what analysts strive to cover, and analysts that stop covering these firms are less likely to be analysts' own decisions (Hong and Kubik, 2003). Moreover, firms in the S&P 500 index tend to be covered by more analysts than other firms that have similar characteristics but are not included in this index (Yu, 2008). Taken together, this sample provides a greater number of potential analyst report transcripts and thus guarantee adequate variation of the analyst report level variable. In addition, during the collection of analyst report transcripts, it seems unavoidable that some of the analyst reports are not included in the database. This could be a more serious issue for collecting the analyst reports for smaller firms that are covered by less analyst. In contrast, to collect the analyst reports for more popular firms are less affected by this problem. Moreover, analysts in this sample are less likely to self-select whether they cover or not to cover the firms in the S&P 500 index, compared to other smaller or unknown firms.

I then download analyst report transcripts for S&P 500 companies during 2015 to 2020 from Investext through Thomson One database. These collected analyst report transcripts are PDF documents that contain the subtitles representing the structure of each report. In addition to the analyst reports, I collect other information from the database that helps identify each report and makes it possible to merge the analyst reports with the data from other databases. The identifier information includes the report name, a unique report id set by the Thomson One database, the release date of the report, the name of the analyst that issue the report, the broker name that employs the analyst, and the number of pages of the transcripts. Such identifier information could work as the id of the report, which is then used to match these analyst reports with the financial data from other databases (e.g., CRSP and

COMPUSTAT).

This sample contains 571,661 analyst report transcripts for 655 S&P 500 companies from 2015 to 2020. The market or industrial research reports are not included in the sample because these reports focus on systematically different content compared to the company research reports. The reports from non-brokers are also not collected. The Thomson One database has the option to directly remove these non-broker providers. For example, several of the excluded providers are the seven “robo-analysts” providers illustrated by Coleman et al. (2021) that use the computer programs to automatically generate research analysis, including the Minkabu, the New Constructs, the Price Target, the Rapid Ratings, the Thestreet.com, the Validea, and the ValuEngine. Other non-broker providers are also removed by this method, which include, for example, the ValuEngine, the Corporate Technology Information Services, and the Marketline - Industry Profiles. I further double-check and remove the unwanted reports according to the procedure below.

After downloading the analyst report transcripts, the several following steps is implemented to select the final sample of analyst reports. Table 1-1 lists the process of selection. First, I delete the duplicate reports that have identical titles with others and are provided by the same analysts for the same firm on the same date. Next, I only retain the analyst reports that are written in English for the S&P 500 companies during the period from 2015 to 2020.

Table 1-1 Sample Selection

This table presents the sample selection procedure for analyst reports. The initial sample includes analyst report transcripts for S&P 500 companies from 2015 to 2020.

Sample selection procedure	Observations
Initial sample of analyst reports of S&P 500 companies from 2015 to 2020	571,661
Less the duplicate reports	-5,361
Less the reports that are not in English	-2,153
Less the reports that are not from brokers	-86,897
Less the reports that are assigned to wrong companies	-4,228
Less the reports that are too short (1th percentile) or too long (more than 40,000 words)	-16,355
Less the reports that have no information about the analysts	-71
Final sample	456,596
Unique brokers	357
Unique firms	612
Average reports per broker in a year	213
Average reports per company in a year	124

Moreover, I manually check the providers of all reports to only retain the analyst reports from brokers and exclude the ones from other unwanted providers or the ones that are clearly not analyst reports[1]. For example, some of the reports from the "robo-analysts" providers are not totally excluded (e.g., the Price Target and the Minkabu) by Thomson One database when downloading. In addition, the downloaded reports from the Thomson Reuters Streetevents are conference call transcripts instead of analyst reports. Some of the reports from the Argus Institutional Partners are market-based analysis rather than company analysis. Similarly, several of the downloaded reports from the Northcoast Research are one-page tables, which contains almost no textual information. Taken together, after manual checking, the downloaded reports from unwanted providers or the ones that are not analyst reports are excluded from the sample.

I then check the analyst reports on each firm and remove the ones that are assigned to the wrong firms[2]. For example, during the checking, a small sample of analyst reports seem to appear in more than one company. Specifically, the analyst reports for the Hess Corporation may contain some reports for the Public Storage, Inc., which are unrelated to the Hess Corporation. Additionally, some of the falsely assigned analyst reports for the Procter & Gamble company are actually reports for the Pacific Gas and Electric Company, in which the name of the Procter & Gamble company is never mentioned. Similarly, some of the analyst reports for the Alphabet Inc. are falsely classified as reports for the Aier Eye hospital group. Thus, these falsely classified reports for these firms are removed, while the analyst reports which are correctly assigned to these companies are contained.

Furthermore, as suggested by Bellstam et al. (2020), I remove analyst reports that have

[1] I exclude all the reports from 33 unwanted providers: the Thomson Reuters Streetevents, the Globaldata, the Pricetarget Research, the Morningstar Credit Research, the Morningstar Credit Thematic Research, the Morningstar Thematic Research, the Sadif Analytics, the Valuengine, Inc, the Valuengine, Inc. , the Validea, the Wright Investors Service, the Minkabu The Inofonoid, Inc. , the Marktfeld, the Buysellsignals Research, the Trefis, the Stock Traders Daily Research, the Corporate Watchdog Reports, the Acquisdata, the Infinata, the Cortellis Company Detailed Drug Pipeline Report, the Cortellis Company And Pipeline Overview Report, the Cortellis Company Competitive Landscape Report, the Zacks Equity Research, the Capitalcube, the Crd Global, the Business Research Company, the Smart Insider, the Tech Charts Llc, the Paragon Intel - Summary Reports, the Responsible Mining Foundation, the Investiger.Net, the Stockbrokers Botswana, the Mcalinden Research Partners.

[2] The excluded analyst reports are from ten companies: CCE-US (the Coca Cola Enterprises Inc New), HES-US (the HESS corporation), PSA-US (the Public Storage incorporated), STI-US (the Suntrust Banks, Inc.), PG-US (the Procter & Gamble company), AGN-US (the Allergan company), BAC-US (the Commerce Bancshares company), USB-US (the Deutsche Bank), F-US (the FDM group holding), and GOOGL-US (the Aier Eye hospital group).

less than 85 words (1th percentile) or higher than 40,000 words. This removes the reports that are extremely short or long to concentrate on a relatively more homogenous set of reports. To set the upper limit of 40,000 words could help keep analysts' initiating reports in the sample, which compose nearly 2% of the whole sample of analyst reports. The calculation of the total number of words in each analyst report transcript is introduced in the next two sections. I finally remove the reports that have no information about their authors (i.e., analysts). The final sample contains 456,596 analyst reports from 357 brokers for 612 companies. On average, one broker has generated 213 reports in a year for these firms, and 124 reports are issued for each firm in a year. This final sample of analyst report transcripts is used in the following sections to conduct the textual analysis.

As described in detail in the Appendix A-3 and Appendix A-4, prior to conducting the textual analysis, several steps are employed to process the downloaded analyst report transcripts. I first utilize the identifier information from the Thomson One database (i.e., the firm's ticker, the release date of the report, the name of the analyst, and the name of the broker) to merge the downloaded analyst reports with the analyst forecast information from the I/B/E/S database. Since the I/B/E/S only provides a 5-digit code to represent each broker and a 6-digit code for each analyst instead of the real name of the broker or analyst, I use the above identifier information to recognize the names of the brokers and analysts that these codes stand for. As a result, I match 161 out of the total 227 brokers to their corresponding broker codes in the I/B/E/S. These 161 brokers account for 362,111 of 456,596 analyst reports in this sample.

Moreover, as elaborated in the Appendix A-4, I exclude the irrelevant information from each of the analyst reports. Such unnecessary elements include tables, graphs, charts, and blank pages, which are not readable texts and do not contain textual information. I also remove the company description, the introduction of broker and analyst, the glossary, the appendix, and the analyst disclosure and disclaimer. These parts do not contain analyst opinions and are irrelevant to the following research, which could distort the outcome of textual analysis. After removing the above unwanted elements, I extract the textual information from each analyst report. Finally, I follow the previous studies to clean and parse the analyst report transcripts (Huang et al., 2014; De Franco et al., 2015; Huang et al., 2018; Bellstam et al., 2020). This procedure includes, for example, converting all words into lower case, transferring some professional financial phrases into individual words, and removing the stop words.

1.4.2. Measures of analyst report similarity

Two measures are employed to capture the textual similarity between the textual content in the analyst report of interest and that in the prior analyst reports, similar to the prior research (Hoberg and Phillips, 2010; Brown and Tucker, 2011; Bozanic and Thevenot, 2015; Hoberg and Phillips, 2016). Specifically, after processing and cleaning all texts following the procedure above, I measure the cosine similarity between the word vectors among analyst reports to construct the main variables. For each analyst report, I select the most recent reports released by all other analysts for the same firm in the 60 days prior to the report.

I only retain the most recent reports by prior analyst to make sure the information set are similar for these prior analysts and the analyst of interest. The previous reports issued by the focal analyst in the prior 60 days are excluded from the calculation. As a robustness test, I also use the period of 30 and 90 days. The results for the following models remain robust.

I use the method below to calculate the textual similarity between an analyst report transcript and the transcripts of other prior analyst reports:

$$Similarity\_raw_i = \frac{1}{N}\sum_{d \in P} \frac{\sum_{j=1}^{n} w_{i,j} w_{d,j}}{\sqrt{\sum_{j=1}^{n} w_{i,j}^2}\sqrt{\sum_{j=1}^{n} w_{d,j}^2}},$$

where P is the time period (i.e., the prior 60 days excluding the date of the focal analyst report), N is the number of analyst reports for the same firm in this period, n is the total number of unique words across all analyst reports for the firm, and $w_{i,j}$ is the frequency of word j as captured by the number of times this word appears is in the focal analyst report i. If a word shows up in the transcript for many times, it would have a higher word frequency for the transcript. Thus, each transcript could be represented by a vector that contains the weight (in this case, the word frequency) of individual words used in the transcript. And this variable, *Similarity_raw*, measures the average cosine similarity between the word vector of the focal analyst report and the word vectors of other corresponding analyst reports released by other analysts for the same firm in the prior 60 days.

The cosine similarity between any two reports ranges from 0 to 1, where 1 indicates perfect similarity and 0 represents no similarity. Specifically, in the case of *Similarity_raw*, a score of 1 suggests that the two reports are identical in terms of their word vectors. This means the two reports contain exactly the same words with identical frequencies. In contrast, a score of 0 implies that the two reports have no similarity in their word vectors, indicating there is no overlap in the words used by the two reports.

The second measure, *Similarity_tfidf*, use a similar way to calculate the textual similarity between an analyst report transcript and the transcripts of other prior analyst reports:

$$Similarity\_tfidf_i = \frac{1}{N}\sum_{d \in P} \frac{\sum_{j=1}^{n} w_{i,j} w_{d,j}}{\sqrt{\sum_{j=1}^{n} w_{i,j}^2}\sqrt{\sum_{j=1}^{n} w_{d,j}^2}},$$

where $w_{i,j}$ is the tf-idf (term frequency–inverse document frequency) weight of word j in the focal analyst report i instead of the raw word frequency. The meanings of other symbols are identical to those in the calculation of *Similarity_raw*. This new measure of the textual similarity conducts the tf-idf (term frequency–inverse document frequency) weighting scheme and use it to substitute for the original weighting method (i.e., the raw number of times a word is used in a transcript). This new weighting approach could capture the importance of a word within a transcript relative to a batch of transcripts. The tf-idf weight of a word in a transcript is the product of the word's term frequency and the inverse document frequency. Particularly, the term frequency is equal to the raw number of times the word appears in the transcript scaled by the total number of words in the transcript. Thus, similar to the prior measure (*Similarity_raw*), a more commonly used word in a transcript has a higher term frequency for the transcript. Furthermore, the inverse document frequency of a word is defined as the natural logarithm value of the total number of transcripts divided by the number of transcripts that contain the word. Therefore, according to this definition, the relatively unique words that appear only in a small number of transcripts are considered more important for these transcripts than other words which are commonly used across transcripts. Taken together, the tf-idf weighting method captures how important a word is within a transcript and across all transcripts.

During the calculation of this variable (*Similarity_tfidf*), the tf-idf weighting scheme are implemented upon the words from analyst report transcripts for S&P 500 companies respectively. That is, according to the calculation of tf-idf weight, the inverse document frequency of a word could vary across firms. To conduct such weighting approach by firm is because the business details are quite different across firms, so do the firm-related words used by analysts in their reports. Some words that are commonly used in the analyst reports for one firm might be rarely mentioned in reports for another firm.

1.4.3. Determinants of analyst report similarity

As a first step, I use the following model to examine whether analysts' tendency to provide similar textual information can be explained by the four factors (i.e., analyst herding behavior, analyst ability, analyst learning behavior, and significant firm-related news) discussed above:

$$
\begin{aligned}
Similarity_{i,j,t} = \alpha &+ \beta_1 Analyst\ herding_{i,j,t} + \beta_2 Analyst\ ability_{i,j,t} \\
&+ \beta_3 Analyst\ learning_{i,j,t} + \beta_4 Firm\ news_{i,j,t} + Controls \\
&+ Firm\ fixed\ effects + Broker\ ixed\ effects \\
&+ Year\ fixed\ effects + \varepsilon_{i,j,t}.
\end{aligned} \tag{1}
$$

The dependent variables (*Similarity_raw* and *Similarity_tfidf*) are the measured textual similarity for the analyst report released for the firm *i* by analyst *j* at time *t*, calculated as the mean value of cosine similarity scores between the word vector of the analyst report and those vestors from the most recent reports issued by other analysts in the prior 60 days. To test the determinants of such textual similarity among analyst reports, I use different variables to proxy for analysts' herding behavior, analyst ability, analysts' learning behavior, and significant firm-related news.

First, I calculate the leader-follower ratio following prior literature (Cooper et al., 2001; Shroff et al., 2014). This leader-follower ratio is calculated as the cumulative number of days by which the two most recent previous reports lead the focal analyst report divided by the cumulative number of days by which the subsequent two reports follow the focal analyst report. The value of this ratio is multiplied by minus 1 so that a higher value indicates that the analyst generating the report is more likely to be a follower analyst than a leader analyst. In particular, the follower analysts tend to be the ones who release their reports immediately after the first mover (leader analysts). In other words, the follower analysts have closer previous reports that can be followed while the prior reports for the leader analysts are in the relatively remote past. This definition is also consistent with the herding definition that a more capable agent acts early and relies on their private information, whereas a less capable agent may take herding behavior in order to hide their low ability (Scharfstein and Stein, 1990; Trueman, 1994). Thus, I assume that the follower analysts are more likely to be the ones who try to follow others (herding).

Next, I add another proxy for analysts' herding behavior. The new variable, *Herding_forecast*, captures analysts' herding behavior when they generate their earnings forecasts for companies, by measuring how greatly an analyst deviates from the consensus opinions among analysts. Specifically, I follow the previous studies to calculate the absolute deviation of an analyst's one year ahead earnings forecast from the most preceding consensus earnings forecast, with this difference scaled by the security price two trading days before the forecast date for the purpose of facilitating comparison across firms (Clement and Tse, 2003; Gleason and Lee, 2003; Clement and Tse, 2005; Jiang et al., 2015; Iselin et al., 2021). Consistent with the dependent variables (*Similarity_raw* and *Similarity_tfidf*), the consensus forecast is measured based on the period of the prior 60 days. That is, the consensus forecast is calculated as the mean value of other analysts' most recent one year ahead earnings forecasts issued in the prior 60 days. The final value of this variable is multiplied by minus 1 so that a higher value indicates less deviation from the prior consensus, indicating that the analyst is more likely to exhibit herding behavior by following others. Overall, I use the analysts' leader and follower status (*Leader-Follower Ratio*) and analysts' deviation from the consensus forecast (*Herding_forecast*) in my model.

Second, I use several variables to capture analysts' ability to collect valuable new information. Specifically, I first calculate the textual similarity between an analyst's report and the most recent prior report from the same analyst for the firm (*Similarity_self*), as captured by the cosine similarity score between the word vector (based on the tf-idf weighting method) of an analyst report and the word vector of the most recent report from the same analyst for the same firm.[3] Therefore, if analysts tend to provide stale textual information in their reports in comparation with their own prior reports, this is likely to suggest that analysts are less capable of providing new information.

In addition, I follow prior literature to calculate the following variables that might capture analyst ability (Stickel, 1995; Mikhail et al., 1997; Clement and Tse, 2005; Bradley et al., 2017; Huang et al., 2017a). Specifically, the *Lag AFE* is defined as the absolute forecast error of the latest one year ahead earnings forecast released by an analyst for the firm in the last year, with the absolute forecast error calculated as the absolute deviation of

[3] The calculation of the main dependent variables (*Similarity_raw* and *Similarity_tfidf*) are not affected by the analysts' own prior reports because these two variables only take into account the prior reports released by other analysts.

an analyst's forecast for the firm from the firm's actual earnings per share in the given year. This absolute difference is scaled by the security price two trading days before the forecast date. This is a reverse measure of analyst forecast accuracy that a higher value indicates less accurate earnings forecasts. The *Broker Size* is defined as the number of analysts employed by the brokerage house that hires the analyst who generates the focal report in the given year. This includes only the analysts from the brokerage who have issued at least one earnings forecast for any firm in a given year. Moreover, the *Forecast Frequency* is defined as the number of one year ahead earnings forecasts issued by an analyst for the firm in a given year. The *Firm Experience* is calculated as the number of years an analyst has generated at least one earnings forecast for the firm. Similarly, the *General Experience* is calculated as the number of years an analyst has generated at least one earnings forecast for any firms. In addition, the *Firms Followed* is calculated as the number of firms covered by an analyst (by issuing at least one earnings forecast for these firms) in a given year. The *Industries Followed* is calculated as the number of two-digit SIC industries followed by an analyst (by issuing at least one earnings forecast for any firms in the industry) in a given year.

Third, to test whether analysts can learn from prior analysts, I explore the methods from the literature (Clement et al., 2011). In their research, the results show that analysts are more likely to use the information contained in the average forecast revision to revise their own forecasts, if the number of analysts included in such average increases or if the accuracy of the average revisions is higher. This is consistent with their argument that analysts have the ability to learn and extract necessary information from other analysts rather than simply herding or free-riding. Therefore, inspired by this method, I calculate two similar variables to examine analysts' possible learning behavior. Specifically, I calculate the number of other analysts' reports contained in the 60 days prior to the date of the focal analyst report (*Number of Reports*). I next calculate the mean value of the absolute forecast errors (*Consensus Accuracy*) for the earnings forecasts issued in the prior 60 days, with the absolute forecast errors scaled by the security price two trading days before the forecast date. The *Consensus Accuracy* is multiplied by minus 1 so that a higher value indicates higher accuracy of the consensus forecast. Hence, it might indicate that analysts have the ability to learn from others and extract useful information, if an analyst report contains more similar textual information with the prior reports when there is a greater number of preceding reports and when the information from these previous reports are more accurate.

Moreover, to capture the effect of firm-related news, I calculate the number of analyst reports issued in the five-day period around the analyst report of interest (*Firm News*).[4] This variable is expected to be positively associated with the important firm-related news around this period (De Franco et al., 2015). Additionally, I include the absolute value of the cumulative ten-day abnormal returns ending the day before the current report date (*Prior |ABR|*), where the abnormal return is calculated as the buy-and-hold return minus the buy-and-hold return on the value-weighted market index for the same period. This variable is used to capture any potential short-term momentum or reversal in stock price because of recent news or events (Huang et al., 2014).

Furthermore, I control for the word counts of each analyst report transcript since the number of words might be correlated with the calculated textual similarity. For example, if two analyst reports have very few words in the transcripts, it is possible that they do not have any common words after the previous processing such as removing the stop words, which could lead to a 0 value of cosine similarity between their word vectors. On the contrary, when each of the two analyst reports contains a great number of unique words, they are more likely to at least share a set of common words. The empirical findings by Brown and Tucker (2011) also indicate that textual similarity tend to be higher when the texts are longer. Hence, such word count is measured for each of the processed transcripts of analyst reports and added to control for such effect. The *Word Counts* is calculated as the natural logarithm of one plus the total number of words in an analyst report transcript.

I also control firm characteristics that might impact information environment. Specifically, the firm size (*Size*) is calculated as the natural logarithm value of the firm's total sales in a given year. The market value (*Market Value*) is calculated as the natural

[4] The five-day period includes the -2, -1, 0, 1, 2 days relative to the report release date.

logarithm value of the firm's market value at the beginning of the year. The book-to-market ratio (*Book-Market ratio*) is calculated as the book value of corporate equity divided by the market value of corporate equity. All continuous variables are winsorized at the 1st and 99th percentiles to mitigate the effect of extreme values. Similar to the prior literature (De Franco et al., 2015), the firm, broker, and year fixed effects are added to control for the time-invariant firm and broker characteristics and the variation across time. Furthermore, each firm could be covered by more than one analyst, while each analyst could follow more than one firm. Therefore, I cluster the standard error at both firm level and analyst level to account for the firm correlation and analyst correlation over time.

1.4.4. Investor reactions to the similarity of analysts' reports

To examine whether the textual similarity among analyst reports could influence the informativeness of analyst reports, I run the following regression:

$$\begin{aligned} Investor\ reaction_{i,j,t} &= \alpha + \beta_1 Similarity_{i,j,t} + Controls + Firm\ fixed\ effects \\ &+ Broker\ fixed\ effects + Year\ fixed\ effects + \varepsilon_{i,j,t}. \end{aligned} \tag{2}$$

The dependent variable is $|ABR|$, calculated as the absolute value of the cumulative three-day abnormal returns starting from the current report date, where the abnormal returns is defined as the buy-and-hold return of the stock minus the buy-and-hold return on the value-weighted market index for the same period. The absolute value of investor reaction is used because the main independent variables are unsigned. In particular, the independent variables are the two measures of textual similarity (*Similarity_raw* and *Similarity_tfidf*), which are defined previously. Ceteris Paribus, if the investor reaction is stronger immediately after the release of an analyst report, this could suggest that the stock abnormal return is likely to be triggered by the analyst report and that this report is informative for investors. Since this is a short-window event study, the observed impact on investor reaction is less likely to be caused by omitted variables and thus this method can help identify a causal relationship (De Franco et al., 2015). Furthermore, I conduct the second stage estimation, which means that *Similarity_raw* and *Similarity_tfidf* in Equation (2) are the predicted values from the estimation of Equation (1).[5]

In addition, I control for several factors that may explain the variation in investor reactions around the release of analyst reports. First, I control for the effect of the tone of analyst reports, because the literature has found that the tone of analyst reports can affect the investor reactions (Asquith et al., 2005; Twedt and Rees, 2012; Huang et al., 2014). Specifically, I obtain the word list of the positive and negative words developed by Loughran and Mcdonald (2011). Similar to the prior study (De Franco et al., 2015), the sentiment score of an analyst report is computed as the number of positive words minus the number of negative words scaled by the total number of positive and negative words in this transcript. The measure of the change in the sentiment of analyst reports, *Tone Rev*, is calculated as the absolute difference between the sentiment score of an analyst's report and the sentiment score of the prior report by the same analyst for the same firm.[6]

To account for the impact of analysts' quantitative outputs released around the report date, I follow the prior studies to control for analysts' revisions of their one year ahead earnings forecasts, their stock recommendations, and their price target forecasts (Huang et al., 2014; De Franco et al., 2015). Specifically, the revisions of analysts' earnings forecasts, *Forecast Rev*, is calculated as the absolute difference between an analyst's earnings forecast issued in the five-day period around the report and the same analyst's prior earnings forecast for the same firm, scaled by the security price two trading days before the forecast date. In addition, the revisions of analysts' stock recommendations, *Recommendation Rev*, is calculated as the absolute difference between an analyst's stock recommendation issued in the five-day report period and the analyst's prior recommendation. The analysts' stock recommendations are numeric values assigned by I/B/E/S database (i.e., 1: Strong Buy; 2:

[5] I also implement an alternative estimation by directly controlling for all determinants in a normal regression and the results are robust.

[6] Additionally, I have tried an alternative measure by directly calculating the absolute value of the sentiment score of an analyst's report. The results using this alternative control variable remain similar.

Buy; 3: Hold; 4: Underperform; 5: Sell). Moreover, the revisions of analysts' price target forecasts, *Target Price Rev*, is calculated as the absolute difference between an analyst's price target forecast issued in the five-day report period and the analyst's prior price target forecast, scaled by the security price two trading days before the forecast date.

Then I control for the timeliness of the report by adding the leader-follower ratio of the report (*Leader-Follower Ratio*). To capture the effect of the recent firm-related news, I control for number of analyst reports issued in the five-day period of the report of interest (*Firm News*) and the absolute value of the cumulative ten-day abnormal returns ending the day before the current report date (*Prior |ABR|*). In addition, the earnings announcement is considered as one of the most important company news. If the firm releases its earnings that is far from investors' expection, the market might react strongly. To control for this effect, I calculate the *|Earnings News|* as the absolute value of earnings announced in the five-day report window minus the most recent analysts' I/B/E/S consensus earnings forecast, scaled by the security price two trading days prior to the earnings announcement. It is set to zero if the firm make no earnings announcement in the five-day report window. Additionally, I control for the number of words in the analyst report transcript (*Word Counts*). I finally control other important firm characteristics using the variables defined above, including the *Size*, the *Market Value*, and the *Book-Market ratio*.

All continuous variables are winsorized at the 1st and 99th percentiles. Similar to the previous model, the firm fixed effect and broker fixed effect are added to control for the time-invariant firm and broker characteristics that might affect investor reaction. The year fixed effect is added to control for the systematic variation across time. Moreover, I use robust standard errors clustered at both analyst and firm level to account for the analyst correlation and firm correlation across time.

1.5. Baseline results

1.5.1. Summary statistics

The Table 1-2 provides the summary statistics of my main variables. The mean (median) value of the first measure of similarity (*Similarity_raw*) is 0.41096 (0.41545). This suggests that, on average, an analyst report has a moderate level of similarity with reports released by other analysts in the preceding 60 days, given that the score of similarity falls within the range of 0 to 1. In contrast, the mean (median) value of the second measure of similarity using tf-idf weighting method (*Similarity_tfidf*) is 0.21641 (0.21562). The difference between the values of two different measures is likely to suggest that the analysts can provide similar textual information in their reports because, to some extent, they use relatively common words, such as “company”, “report”, “earnings”, and “EPS”, which may appear in many of the other analyst reports. Consequently, after adjusting for the weight of these words using the tf-idf weighting method, the calculated average of cosine similarity scores between the word vectors among analyst reports tends to be lower.

Table 1-2 Summary Statistics

This table presents the summary statistics results, including the number of observations, mean, median, standard deviation, first quartile, and third quartile of variables used in this empirical analysis. For definitions of these variables, please refer to Table A-1. All continuous variables are winsorized at the 1st and 99th percentiles.

Variable	N	Mean	Median	Std. dev	Q1	Q3
Similarity_raw	124,515	0.41096	0.41545	0.07580	0.36135	0.46458
Similarity_tfidf	124,515	0.21641	0.21562	0.05605	0.17753	0.25445
Word Counts Raw	124,515	1417.67	1072.00	1188.63	671.00	1716.00
Word Counts	124,515	7.00148	6.97821	0.69827	6.51026	7.44833
Negative	124,515	0.01364	0.01238	0.00729	0.00842	0.01752
Positive	124,515	0.01388	0.01301	0.00702	0.00883	0.01811
Determinants						
Leader-Follower Ratio	124,515	-2.35616	-1.00000	3.67614	-2.48000	-0.42857
Herding_forecast	124,515	-0.00402	-0.00116	0.00904	-0.00336	-0.00039
Similarity_self	124,515	0.55545	0.54938	0.22642	0.37893	0.72521
Lag AFE	124,515	0.00252	0.00086	0.00518	0.00033	0.00225
Broker Size	124,515	85.44412	74.00000	59.48556	31.00000	123.00000
Forecast Frequency	124,515	6.24252	6.00000	2.69032	4.00000	8.00000
Firm Experience	124,515	7.39013	6.00000	4.93159	3.00000	10.00000
General Experience	124,515	14.23469	14.00000	7.95655	8.00000	20.00000
Firms Followed	124,515	23.00382	22.00000	8.92197	17.00000	27.00000
Industries Followed	124,515	3.34222	3.00000	1.55571	2.00000	4.00000
Number of Reports	124,515	11.32253	11.00000	5.93873	7.00000	15.00000
Consensus Accuracy	124,515	-0.00874	-0.00324	0.01681	-0.00805	-0.00130
Firm News	124,515	15.47583	16.00000	9.61456	7.00000	23.00000
Prior |ABR|	124,515	0.04046	0.02852	0.04036	0.01289	0.05347
Other variables						
|ABR|	124,515	0.03772	0.02495	0.03923	0.01087	0.05023
ABR	124,515	0.00082	0.00072	0.05208	-0.02413	0.02578
Volume	124,515	0.44270	0.39815	0.45881	0.11702	0.71959
Tone Rev	124,515	0.28933	0.21990	0.25560	0.09524	0.40988
Forecast Rev	124,515	0.00318	0.00102	0.00558	0.00025	0.00316
Recommendation Rev	124,515	0.02371	0.00000	0.15214	0.00000	0.00000
Target Price Rev	124,515	0.04356	0.00000	0.07208	0.00000	0.06295
|Earnings News|	124,515	0.00124	0.00033	0.00256	0.00000	0.00125
Size	124,515	9.37799	9.27191	1.22737	8.49941	10.08664
Market Value	124,515	10.11155	9.96586	1.18367	9.27354	10.87054
Book-Market Ratio	124,515	0.42933	0.31885	0.36515	0.17391	0.57423
Earnings Bad News	124,515	-0.00061	0.00000	0.00217	-0.00092	0.00000
Uncertainty	124,515	0.01424	0.01344	0.00642	0.00958	0.01807

On average, the transcript of an analyst report for the S&P 500 company contains 1417.67 words after the processing (e.g., removing stop words and tables). The percentage of negative words in an analyst report transcript is 1.36%, while the percentage of positive words is 1.39%. This may suggest that analysts covering the S&P 500 companies are overall more optimistic, but only to very little extent. Furthermore, each analyst report is accompanied by 11.32 previous reports issued by 11.32 unique analysts in the prior 60 days. In addition, the analysts covering the S&P 500 firms have worked as sell-side analysts for 14.23 years and have issued forecasts for each company for 7.39 years. These analysts simultaneously cover 20.00 firms from 3.34 industries in a year. And they release 6.24 one year ahead earnings forecasts for each firm in a year.

The statistical properties (e.g., the mean, median, and standard deviation) of these variables are mostly consistent with the prior studies. For example, the mean (median) value of the measure of investor reaction (|*ABR*|) is 0.03772 (0.02495), while the mean (median) value of the raw measure (*ABR*) is 0.00082 (0.00072). This is consistent with the values reported by Huang et al. (2014).

The pearson correlations between the variables for the determinants and consequences tests are presented in Table 1-3. The Panel A reports the correlation coefficients between variables used in Equation (1), while Panel B lists the correlation coefficients between variables used in Equation (2). The results show that the two measures of textual similarity (*Similarity_raw* and *Similarity_tfidf*) are highly correlated with each other. Consistent to the expectation, the number of words in an analyst report is positively associated with the two measures of textual similarity. In addition, the investor reaction (|*ABR*|) is positivelly correlated with analysts' revisions of their earnings forecasts and target price forecasts. However, the coefficients between most variables are relatively low.

Table 1-3 Correlation Matrix

This table shows correlations between variables. Variable definitions are provided in Appendix A. * indicates statistical significance at the 1% level.

Panel A: Correlations between determinants of analyst report similarity

Variables	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)	(11)	(12)	(13)	(14)	(15)	(16)	(17)	(18)	(19)	(20)
(1) Similarity_raw	1.00																			
(2) Similarity_tfidf	0.92*	1.00																		
(3) Leader-Follower Ratio	0.10*	0.09*	1.00																	
(4) Herding_forecast	0.04*	0.05*	0.03*	1.00																
(5) Similarity_self	0.18*	0.18*	0.01*	-0.01*	1.00															
(6) Lag AFE	-0.03*	-0.03*	0.00	-0.38*	0.03*	1.00														
(7) Broker Size	-0.01*	-0.00	-0.03*	-0.01*	-0.08*	0.01	1.00													
(8) Forecast Frequency	-0.10*	-0.11*	-0.02*	-0.14*	-0.11*	0.06*	0.14*	1.00												
(9) Firm Experience	0.01*	0.03*	-0.02*	0.00	-0.03*	-0.03*	-0.03*	0.04*	1.00											
(10) General Experience	0.07*	0.08*	-0.01	0.00	-0.02*	-0.03*	-0.07*	0.00	0.58*	1.00										
(11) Firms Followed	0.04*	0.04*	-0.01*	-0.04*	0.16*	0.04*	0.03*	-0.02*	0.03*	0.15*	1.00									
(12) Industries Followed	0.11*	0.09*	0.04*	0.04*	0.16*	-0.03*	-0.13*	-0.18*	-0.02*	0.08*	0.37*	1.00								
(13) Number of Reports	-0.01*	-0.03*	0.36*	0.01*	-0.02*	-0.01*	-0.07*	0.04*	-0.01*	0.00	0.05*	0.05*	1.00							
(14) Consensus Accuracy	0.03*	0.03*	0.02*	0.67*	-0.01	-0.36*	-0.01	-0.17*	-0.01*	-0.02*	-0.05*	0.01*	0.01*	1.00						
(15) Firm News	-0.05*	-0.05*	0.01	0.07*	-0.05*	-0.10*	-0.01	-0.09*	0.03*	0.02*	-0.02*	0.01*	0.16*	0.10*	1.00					
(16) Prior |ABR|	0.01*	0.01*	0.08*	-0.25*	0.01*	0.15*	-0.03*	0.07*	-0.04*	-0.02*	0.03*	0.01*	0.11*	-0.27*	-0.07*	1.00				
(17) Word Counts	0.32*	0.37*	0.07*	-0.01*	0.29*	0.02*	0.06*	-0.09*	-0.04*	-0.06*	0.12*	0.10*	0.09*	-0.01*	-0.11*	0.04*	1.00			
(18) Size	-0.12*	-0.10*	0.06*	0.07*	-0.05*	-0.07*	0.04*	0.07*	0.12*	0.08*	-0.04*	0.06*	0.29*	0.06*	0.19*	-0.11*	0.06*	1.00		
(19) Market Value	-0.10*	-0.09*	0.09*	0.15*	-0.05*	-0.17*	0.03*	0.07*	0.09*	0.08*	0.02*	0.03*	0.40*	0.18*	0.26*	-0.17*	0.05*	0.73*	1.00	
(20) Book-Market Ratio	-0.01*	0.01*	-0.04*	-0.28*	0.01*	0.34*	0.02*	0.20*	0.03*	0.07*	0.03*	-0.14*	-0.08*	-0.33*	-0.13*	0.10*	-0.01*	0.03*	-0.26*	1.00

Panel B: Correlations between independent variables in investor reactions															
Variables	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)	(11)	(12)	(13)	(14)	(15)
(1) |ABR|	1.00														
(2) Similarity_raw	-0.04*	1.00													
(3) Similarity_tfidf	-0.06*	0.92*	1.00												
(4) Tone Rev	0.02*	-0.12*	-0.13*	1.00											
(5) Forecast Rev	0.19*	-0.04*	-0.05*	-0.01	1.00										
(6) Recommendation Rev	0.01*	0.01*	0.01*	0.02*	0.02*	1.00									
(7) Target Price Rev	0.19*	0.01*	0.01	0.05*	0.18*	0.17*	1.00								
(8) Leader-Follower Ratio	-0.05*	0.10*	0.09*	-0.02*	-0.03*	0.02*	0.02*	1.00							
(9) Firm News	0.18*	-0.05*	-0.05*	0.01*	-0.06*	-0.09*	0.00	0.01	1.00						
(10) Prior |ABR|	0.11*	0.01*	0.01*	-0.01*	0.24*	0.05*	0.20*	0.08*	-0.07*	1.00					
(11) |Earnings News|	0.19*	-0.03*	-0.03*	-0.00	0.43*	-0.03*	0.10*	-0.07*	0.22*	0.12*	1.00				
(12) Word Counts	-0.05*	0.32*	0.37*	-0.24*	0.01*	0.05*	-0.04*	0.07*	-0.11*	0.04*	-0.04*	1.00			
(13) Size	-0.12*	-0.12*	-0.10*	0.00	-0.04*	-0.02*	-0.07*	0.06*	0.19*	-0.11*	-0.01*	0.06*	1.00		
(14) Market Value	-0.17*	-0.10*	-0.09*	0.01*	-0.15*	-0.02*	-0.10*	0.09*	0.26*	-0.17*	-0.13*	0.05*	0.73*	1.00	
(15) Book-Market Ratio	0.01*	-0.01*	0.01*	0.01	0.26*	0.01*	0.02*	-0.04*	-0.13*	0.10*	0.22*	-0.01*	0.03*	-0.26*	1.00

1.5.2. Determinants of analyst report similarity

In this section, I estimate the regression model in Equation (1) to examine whether the analysts' tendency to provide similar textual information in their reports can be explained by the four determinants previously discussed. Table 1-4 presents the corresponding results. The dependent variables are the measured textual similarity for each analyst report (*Similarity_raw* and *Similarity_tfidf*). The proxies for analysts' herding behavior are the *Leader-Follower Ratio* and the *Herding_forecast*. The analysts' ability is captured by the *Similarity_self, Lag AFE, Broker Size, Forecast Frequency, Firm Experience, General Experience, Firms Followed,* and *Industries Followed*. In addition, to investigate analysts' learnings bahavior, two variables (*Number of Reports* and *Consensus Accuracy*) are included in the regression. Finally, the significant firm-related news around the analyst report is proxied by *Firm News* and *Prior |ABR|*. Other control variables include *Word Counts, Size, Market Value,* and *Book-Market ratio*. The definitions of these variables are discussed in Table A-1.

Table 1-4 Determinants of Analyst Report Similarity

This table presents the results of regressions of analysts' tendency to provide similar textual information in their reports on several determinants and other control variables. The dependent variables are *Similarity_raw*, defined as the mean value of cosine similarity scores between the word frequency vector of an analyst report and those from the most recent reports issued by other analysts in the prior 60 days; *Similarity_tfidf*, defined as the mean value of cosine similarity scores between the word frequency vector of an analyst report and those from the most recent reports issued by other analysts in the prior 60 days, where the word frequency vector is weighted adjusted using term frequency–inverse document frequency (tf-idf) weighting method. The independent variables include *Leader-Follower Ratio, Herding_forecast, Similarity_self, Lag AFE, Broker Size, Forecast Frequency, Firm Experience, General Experience, Firms Followed, Industries Followed, Number of Reports, Consensus Accuracy, Firm News, Prior |ABR|, Word Counts, Size, Market Value*, and *Book-Market Ratio*. For definitions of these variables, please refer to Table A-1. All continuous variables are winsorized at the 1st and 99th percentiles. The firm, broker, and year fixed effects are added. The standard error is clustered at both the analyst and firm level. T-statistics are reported in parentheses. *, **, and *** indicate statistical significance at the 10%, 5%, and 1% levels, respectively.

	(1)	(2)
	Similarity_raw	Similarity_tfidf
Leader-Follower Ratio	0.001***	0.001***
	(12.485)	(13.453)
Herding_forecast	0.123***	0.133***
	(2.923)	(4.315)
Similarity_self	0.029***	0.021***
	(10.752)	(10.004)
Lag AFE	-0.077	-0.071
	(-0.915)	(-1.168)
Broker Size	0.000	0.000
	(0.349)	(0.329)
Forecast Frequency	-0.001***	-0.001***
	(-5.729)	(-5.054)
Firm Experience	-0.000	-0.000
	(-1.365)	(-0.178)
General Experience	0.000**	0.000*
	(2.343)	(1.808)
Firms Followed	-0.000	-0.000
	(-1.104)	(-0.005)
Industries Followed	0.000	0.000
	(0.364)	(0.013)
Number of Reports	0.001***	0.001***
	(9.034)	(7.466)
Consensus Accuracy	-0.049	-0.034
	(-1.564)	(-1.558)

Firm News	0.000***	0.000***
	(6.436)	(7.694)
Prior |ABR|	0.041***	0.031***
	(5.592)	(5.909)
Word Counts	0.045***	0.038***
	(41.581)	(45.260)
Size	-0.002	-0.002
	(-0.624)	(-0.975)
Market Value	-0.006**	-0.004***
	(-2.395)	(-2.679)
Book-Market Ratio	-0.006	-0.003
	(-1.637)	(-1.226)
Firm fixed effects	Yes	Yes
Broker fixed effects	Yes	Yes
Year fixed effects	Yes	Yes
Number of observations	124515	124515
Adjusted R-squared	0.420	0.426

The results are mostly consistent with the idea that analysts tend to provide similar textual information in their reports when they are likely to have herding behavior, when they are less capable of collecting valuable new information, when they are trying to learn from others, and when there is significant firm-related news around the report date. Specifically, the coefficients on *Leader-Follower Ratio* in Columns 1 and 2 (0.001 and 0.001) are significantly positive at the 1% level. This indicates that analysts who are more likely to be identified as follower analysts tend to provide textual information that is more similar to the prior analysts' reports. On the contrary, the leader analysts or the first-mover analysts are less likely to provide similar textual information. Besides, the coefficients on *Herding_forecast* (0.123 and 0.133) are significantly positive, suggesting that when analysts have a tendency of taking herding behavior as reflected in their earnings forecasts, these analysts are more likely to provide similar textual information in their reports.[7] Taken together, these results indicate that analysts tend to provide similar textual information in reports when they tend to have herding behavior.

In addition, I observe significantly positive coefficients on *Similarity_self* in Columns 1 and 2 (0.029 and 0.021). This indicates that if analysts are less capable of collecting and providing new information in their reports compared to their most recent prior reports, they are more likely to offer the stale information that has already been discussed in other previous analysts' reports. Moreover, the results show that the coefficients on *Forecast Frequency* (-0.001 and -0.001) are significantly negative at the 1% level. This suggests that analysts with higher ability (by revising their forecasts more frequently) are less likely to provide similar textual information. However, it is observed that the coefficients on *General Experience* (0.000 and 0.000) are significantly positive at the 5% and 10% level, respectively This could be possibly explained by the incentive argument that younger analysts with less work experience may have stronger incentives to deviate themselves from the consensus, for the purpose of generating publicity (Soltes, 2014; Frijns and Huynh, 2018). The coefficients are not statistically significant for other variables (i.e., *Lag AFE, Broker Size, Firm*

[7] An alternative measure that captures similarity among analysts' quantitative metrics is the forecast dispersion, calculated as the standard deviation of analysts' earnings forecasts in the prior 60 days, scaled by the two-day lagged security price. The regression results hold after adding this variable as a new control.

Experience, Firms Followed, and *Industries Followed*) that proxy for analysts' ability. Overall, although these results are likely to suggest that analysts' ability could play a role in the analyst report similarity, the observed association does not seem to be consistently significant.

Furthermore, I find that the coefficients on *Number of Reports* (0.001 and 0.001) are significantly positive. This suggests that when there is a greater number of reports from other analysts issued in the prior 60 days, an analyst is more likely to provide textual information that is more similar to these prior reports.[8] I do not observe significant coefficients on *Consensus Accuracy*.[9] Taken together, these results likely indicate that analysts are likely to have learning behavior by using the necessary and relevant information provided by prior analysts.

Finally, I detect significantly positive coefficients on *Firm News* (0.000 and 0.000) and *Prior |ABR|* (0.041 and 0.031). This indicates that an analyst is more likely to provide textual information that is similar to prior analyst reports, when there is significant firm-related news as captured by the number of analyst reports around the short-term period and the stock's absolute abnormal return immediately before the report date. These results are mostly consistent with that when there is significant firm-related news around the analyst reports, analysts are more likely to discuss and analyze the information about the news, leading to more similar textual information across their reports. Overall, these results are mostly consistent with the hypothesis H1.[10]

---

[8] As a robustness test, I employ an alternative measure: the number of analysts issuing earnings forecasts in the 60 days prior to the focal analyst report. The results remain consistent with the original findings.

[9] However, in an unreported result, I add an interaction term between *Number of Reports* and *Consensus Accuracy*. I observe significantly positive coefficients on the interaction term, suggesting that analysts tend to generate textual information that is more similar to previous reports from prior analysts, if the consensus opinions among these prior analysts are more accurate.

[10] Moreover, as suggested by prior research, the corporate earnings announcement (EA) is considered by analysts to be among the most significant firm-related news events, with a large portion of analyst reports and forecasts issued immediately following this event (Livnat and Zhang, 2012; Huang et al., 2018). As a robustness test, I examine whether the observed original results hold when analyst reports are divided into two groups based on whether they are released within the four-day period immediately after the EA date. The results for these two groups are mostly consistent with the original results, except for the coefficients on *Firm News*, which seem to be negative for the sample of reports issued right after the EA. This is possibly because, in the short period immediately after EA, the variation in this variable likely captures factors other than the presence of significant firm-related news, given that the sample group is already focused on the most significant firm-related news event (i.e., EA).

With respect to the control variables, the results show that the coefficients on *Word Counts* are significantly positive at the 1% level in Column 1 and 2, consistent with the prediction that textual similarity tend to be higher when the texts are longer. Moreover, I observe significantly negative coefficients on *Market Value*, one of the measures of the complexity of corporate information environment. The negative coefficients indicate that analysts are less likely to provide similar textual information if they work in a more complex information environment. Additionally, no significant coefficient is found for the remaining firm controls (*Size* and *Book-Market ratio*).

Overall, these empirical evidences are mostly consistent with that analysts tend to provide similar textual information in their reports when they are likely to take herding behavior, when they are less capable of collecting valuable new information, when they are learning from others, and when there is significant firm-related news.

1.5.3. Investor reactions to the similarity

In this section, I investigate the impact of the similarity among the textual information in analyst reports on the informativeness of these reports using Equation (2). Table 1-5 reports the estimating results of regressions of investor reaction on analyst report similarity and other control variables. The dependent variable is |*ABR*|, calculated as the absolute three-day abnormal stock return starting from the current report date. The independent variables are the predicted values of *Similarity_raw* and *Similarity_tfidf* based on the previous estimation of the regression in Equation (1). The controls include *Tone Rev*, *Forecast Rev*, *Recommendation Rev*, *Target Price Rev*, *Leader-Follower Ratio*, *Firm News*, *Prior* |*ABR*|, |*Earnings News*|, *Word Counts*, *Size, Market Value,* and *Book-Market ratio*. The detailed definitions of these variables are provided in Table A-1.

Table 1-5 Investor Reactions to Analyst Report Similarity

This table presents the results of regressions of investor reactions on analysts' tendency to provide similar textual information in their reports. The dependent variable is |*ABR*|, defined as the absolute value of the cumulative three-day abnormal returns starting from the current report date, with abnormal returns calculated as the buy-and-hold return minus the buy-and-hold return on the value-weighted market index for the same period. The independent variables are *Similarity_raw*, defined as the mean value of cosine similarity scores between the word frequency vector of an analyst report and those from the most recent reports issued by other analysts in the prior 60 days; *Similarity_tfidf*, defined as the mean value of cosine similarity scores between the tf-idf weighted word vector of an analyst report and tf-idf weighted word vectors from the most recent reports issued by other analysts in the prior 60 days. These two independent variables are the predicted values of *Similarity_raw* and *Similarity_tfidf* based on the estimation of the regression in Equation (1). Other explanatory variables include *Tone Rev, Forecast Rev, Recommendation Rev, Target Price Rev, Leader-Follower Ratio, Firm News, Prior |ABR|, |Earnings News|, Word Counts, Size, Market Value*, and *Book-Market Ratio*. For definitions of these variables, please refer to Table A-1. All continuous variables are winsorized at the 1st and 99th percentiles. The firm, broker, and year fixed effects are added. The standard error is clustered at both the analyst and firm level. T-statistics are reported in parentheses. *, **, and *** indicate statistical significance at the 10%, 5%, and 1% levels, respectively.

	(1)	(2)
	|ABR|	|ABR|
Similarity_raw	-0.212***	
	(-8.792)	
Similarity_tfidf		-0.297***
		(-9.022)
Tone Rev	0.001***	0.001***
	(2.999)	(2.988)
Forecast Rev	0.592***	0.584***
	(8.639)	(8.545)
Recommendation Rev	0.002***	0.002***
	(2.624)	(2.610)
Target Price Rev	0.059***	0.059***
	(14.218)	(14.212)
Leader-Follower Ratio	-0.000***	-0.000***
	(-3.617)	(-3.443)
Firm News	0.001***	0.001***
	(17.864)	(18.166)
Prior |ABR|	-0.007	-0.007
	(-0.690)	(-0.704)
|Earnings News|	0.411**	0.410**
	(2.363)	(2.359)
Word Counts	0.008***	0.009***
	(6.694)	(7.190)

Size	-0.007***	-0.007***
	(-3.111)	(-3.198)
Market Value	-0.003*	-0.003*
	(-1.718)	(-1.774)
Book-Market Ratio	0.009***	0.010***
	(3.439)	(3.539)
Firm fixed effects	Yes	Yes
Broker fixed effects	Yes	Yes
Year fixed effects	Yes	Yes
Number of observations	124515	124515
Adjusted R-squared	0.246	0.246

The results in Table 1-5 show that the value of the adjusted R-square is 24.6% in both Column 1 and 2, which is comparable with prior literature (De Franco et al., 2015; Huang et al., 2018). The coefficients on *Similarity_raw* (-0.212) and *Similarity_tfidf* (-0.297) are significantly positive at the 1% level. This negative association between the analyst report similarity and the investor reaction suggests that such textual similarity among analyst reports could reduce the informativeness of analyst reports. These results are mostly consistent with the hypothesis H2a that an analyst report is less informative for market investors if the textual information in the report is similar to that in prior reports.

With respect to the control variables, it is observed that the coefficients on analysts' revision of the tone of their reports (*Tone Rev*) are significantly positive, suggesting that the investor reaction is higher when the tone of the textual information in analyst reports has changed. Similarly, I observe significantly positive coefficients on *Forecast Rev*, *Recommendation Rev*, and *Target Price Rev*, which indicates that the investor reaction is increased when analysts have higher revisions in their earnings forecasts, stock recommendations, and price target forecasts. These results are also consistent with the findings of Huang et al. (2014) that the both qualitative and quantitative information in analyst reports could provide value to investors. In addition, the significantly positive coefficients on *Firm News* and |*Earnings News*| indicates that the significant firm-related news is associated with higher investor reaction.

As a robustness test, I use an alternative regression model to examine the association between the similarity among the textual information in analyst reports and the informativeness of these reports. I repeat the estimation of the regression model in Equation (2) by using the original value of the two measures of textual similarity (*Similarity_raw* and *Similarity_tfidf*) rather than the predicted value from the model in Equation (1). Furthermore, I include the independent variables in Equation (1) as control variables and add them to the regression model in Equation (2). The estimation results of this model is reported in Table A-2 in the Appendix. The results show that the coefficients on *Similarity_raw* (-0.015) and *Similarity_tfidf* (-0.024) are positive and significant at the 1% level, after controlling for this set of analyst and firm characteristics. This is consistent with the baseline results in Table 1-5.

Furthermore, to see whether this relationship can hold for a long-term period, I repeat the regression Equation (2) by replacing the dependent variable (|*ABR*|) with the absolute abnormal return calculated based on different time horizons. Specifically, the absolute abnormal return is calculated based on 0-, 1-, 2-, 21-, 42-, and 126-day period after the report date (i.e., the day 0). For example, the |*ABR_21*| is calculated as the absolute value of abnormal stock return for the 21 trading days (nearly one calendar month) after the report date (with the report date included). All other settings are identical with Equation (2).

Table 1-6 presents the results of the regression based on the new dependent variables (i.e., the absolute abnormal return based on different time horizons). The results show consistent negative coefficients on *Similarity_raw* in Panel A and *Similarity_tfidf* in Panel B. This negative association between the measures of analyst report similarity and investor reaction remain statistically significant for a relatively long-term period. That is, the analyst report similarity is associated with reduced absolute abnormal stock return for a period of at least 42 trading days in Panel A and 126 trading days in Panel B. Taken together, the results in Table 1-5 and Table 1-6 are consistent with the hypothesis H2a that analyst reports are less informative for market investors if analysts tend to provide similar textual information in reports.

Table 1-6 Investor Reactions based on Different Time Period

This table presents the results of regressions of investor reactions on analysts' tendency to provide similar textual information in their reports, by calculating |*ABR*| based on different time horizons. Specifically, the absolute abnormal return is calculated based on 0-, 1-, 2-, 21-, 42-, and 126-day period after the report date (i.e., the day 0). Other explanatory variables include *Tone Rev, Forecast Rev, Recommendation Rev, Target Price Rev, Leader-Follower Ratio, Firm News, Prior |ABR|, |Earnings News|, Word Counts, Size, Market Value*, and *Book-Market Ratio*. For definitions of these variables, please refer to Table A-1. All continuous variables are winsorized at the 1st and 99th percentiles. For brevity, the coefficients on control variables are not reported. The firm, broker, and year fixed effects are added. The standard error is clustered at both the analyst and firm level. T-statistics are reported in parentheses. *, **, and *** indicate statistical significance at the 10%, 5%, and 1% levels, respectively.

Panel A: Similarity_raw						
	(1)	(2)	(3)	(4)	(5)	(6)
	|ABR_0|	|ABR_1|	|ABR_2|	|ABR_21|	|ABR_42|	|ABR_126|
Similarity_raw	-0.181***	-0.203***	-0.212***	-0.297***	-0.217***	-0.111
	(-8.235)	(-8.716)	(-8.792)	(-8.600)	(-5.186)	(-1.397)
Controls	Yes	Yes	Yes	Yes	Yes	Yes
Firm fixed effects	Yes	Yes	Yes	Yes	Yes	Yes
Broker fixed effects	Yes	Yes	Yes	Yes	Yes	Yes
Year fixed effects	Yes	Yes	Yes	Yes	Yes	Yes
Number of observations	124515	124515	124515	124515	124515	124515
Adjusted R-squared	0.227	0.253	0.246	0.224	0.240	0.308

Panel B: Similarity_tfidf						
	(1)	(2)	(3)	(4)	(5)	(6)
	|ABR_0|	|ABR_1|	|ABR_2|	|ABR_21|	|ABR_42|	|ABR_126|
Similarity_tfidf	-0.243***	-0.277***	-0.297***	-0.439***	-0.344***	-0.219**
	(-8.034)	(-8.783)	(-9.022)	(-9.365)	(-6.009)	(-2.051)
Controls	Yes	Yes	Yes	Yes	Yes	Yes
Firm fixed effects	Yes	Yes	Yes	Yes	Yes	Yes
Broker fixed effects	Yes	Yes	Yes	Yes	Yes	Yes
Year fixed effects	Yes	Yes	Yes	Yes	Yes	Yes
Number of observations	124515	124515	124515	124515	124515	124515
Adjusted R-squared	0.227	0.252	0.246	0.224	0.241	0.308

1.6. Further results

1.6.1. Interaction with analysts' quantitative outputs

Prior literature suggests that the textual information in analyst reports can provide extra value to investors by helping investors understand the information contained in the quantitative information provided by analysts (Huang et al., 2014). To contribute to the current research, I study whether higher similarity among the textual information in analyst reports can affect the informativeness of analysts' quantitative outputs. Following prior literature (Huang et al., 2014; De Franco et al., 2015), I use the three control variables (*Forecast Rev*, *Recommendation Rev*, and *Target Price Rev*) that capture three important quantitative outputs provided by analysts: the earnings forecasts, the stock recommendations, and the price target forecasts. Specifically, I define *D_Forecast Rev* as equal to one if *Forecast Rev* is higher than the sample median in a given year and zero otherwise. Similarly, I define *D_Recommendation Rev* as equal to one if *Recommendation Rev* is higher than the sample median in a given year and zero otherwise, and *D_Target Price Rev* as equal to 1 if *Target Price Rev* is higher than the sample median in a given year and zero otherwise. I expand Equation (2) by including interaction terms between the two similarity measures (*Similarity_raw* and *Similarity_tfidf*) and each of the measures of analysts' quantitative outputs (*D_Forecast Revision*, *D_Recommendation Rev*, and *D_Target Price Rev*). The three continuous measure of these quantitative outputs (*Forecast Rev*, *Recommendation Rev*, and *Target Price Rev*) are excluded from the regression due to the correlation between them and the three dummy variables.

The results are presented in Table 1-7. The coefficients on the interaction terms between the two measures of analyst report similarity and analysts' revisions of earnings forecasts (-0.016 and -0.019) or stock recommendations (-0.070 and -0.087) are significantly negative. Similarly, I observe negative coefficients on the interaction terms between the two measures of analyst report similarity and analysts' revisions of their target price forecasts (-0.010 and

-0.012), though these coefficients are not significant. These results show that the analysts' quantitative outputs are less informative for investors if analysts tend to provide similar textual information in their reports. This is consistent with that analysts' textual information can provide value to investors by helping them understand the quantitative outputs (Huang et al., 2014). If analysts cannot provide new textual information in their reports to justify their quantitative outputs, investors may perceive such quantitative outputs as less informative.

Table 1-7 Interaction with Analysts' Quantitative Outputs

This table presents the results of regressions of investor reactions on analyst report similarity, and its interaction with analysts' revisions of earnings forecasts, stock recommendations, and price target forecasts. The dependent variable is *|ABR|*. The independent variables are *Similarity_raw* and *Similarity_tfidf*. *D_Forecast Rev* equals one if the *Forecast Rev* of the analyst is higher than the sample median in a given year and zero otherwise. *D_Recommendation Rev* equals one if the *Recommendation Rev* of the analyst is higher than the sample median in a given year and zero otherwise. *D_Target Price Rev* equals one if the *Target Price Rev* of the analyst is higher than the sample median in a given year and zero otherwise. Other explanatory variables include *Tone Rev, Forecast Rev, Recommendation Rev, Target Price Rev, Leader-Follower Ratio, Firm News, Prior |ABR|, |Earnings News|, Word Counts, Size, Market Value*, and *Book-Market Ratio*. For definitions of these variables, please refer to Table A-1. All continuous variables are winsorized at the 1st and 99th percentiles. For brevity, the coefficients on control variables are not reported. The firm, broker, and year fixed effects are added. The standard error is clustered at both the analyst and firm level. T-statistics are reported in parentheses. *, **, and *** indicate statistical significance at the 10%, 5%, and 1% levels, respectively.

	(1)	(2)
	|ABR|	|ABR|
Similarity_raw	-0.195***	
	(-7.786)	
Similarity_raw*D_Forecast Rev	-0.016**	
	(-1.984)	
Similarity_raw*D_Recommendation Rev	-0.070***	
	(-3.139)	
Similarity_raw*D_Target Price Rev	-0.010	
	(-1.135)	
Similarity_tfidf		-0.290***
		(-8.432)
Similarity_tfidf*D_Forecast Rev		-0.019*
		(-1.916)
Similarity_tfidf*D_Recommendation Rev		-0.087***
		(-3.158)
Similarity_tfidf*D_Target Price Rev		-0.012
		(-1.140)
D_Forecast Rev	0.011***	0.009***
	(3.256)	(3.837)
D_Recommendation Rev	0.035***	0.025***
	(3.623)	(3.879)
D_Target Price Rev	0.008**	0.007***
	(2.322)	(2.994)
Controls	Yes	Yes
Firm fixed effects	Yes	Yes

Broker fixed effects	Yes	Yes
Year fixed effects	Yes	Yes
Number of observations	124515	124515
Adjusted R-squared	0.236	0.236

1.6.2. Interaction with firms' bad news

Prior studies indicate that when the firm managers have bad news to deliver, they may have a tendency to hide relevant information because this could decrease their human capital and reputation (Lang and Lundholm, 1993; Miller, 2002; Chen et al., 2011). Hence, it is reasonable to expect that under this condition, investors could place a greater value on analysts' efforts to provide new information (e.g., the relevant corporate information that is hidden by company leaders). Thus, I expect the negative relationship between investors' reaction and textual similarity among analyst reports is more pronounced when there is bad news about the firm. To examine this effect, I measure the distance between a firm's earnings announced in the five-day report window and the most recent analysts' I/B/E/S consensus earnings forecast, scaled by the security price two trading days prior to the earnings announcement (*Earnings Bad News*). The *Earnings Bad News* is set to 0 if the firm make no earnings announcement in the five-day report window. The *Earnings Bad News* is multiplied by minus 1. Therefore, a positive value means that the actual corporate earnings is lower than analysts' expectation, while the negative value indicates a firm's earnings larger than analysts' prediction.

Specifically, the *D_Earnings Bad News* is defined as equal to one if the *Earnings Bad News* is higher than the sample median in a given year and zero otherwise. I expand Equation (2) by including the interaction terms between each of the two similarity measures (*Similarity_raw* and *Similarity_tfidf*) and *D_ Earnings Bad News*. Table 1-8 below lists the results. It could be observed that the coefficients on the interaction terms (-0.024 and -0.028) are negative and significant in Column 1 and 2. This suggests that investors are likely to put higher value on analysts' efforts to provide relevant new information when the firm managers have more incentives to hide information.

Table 1-8 Further Results

This table presents the results of regressions of investor reactions on analyst report similarity, and its interaction with firms' bad news, firm uncertainty, brokerage size, and analyst experience. The dependent variable is |*ABR*|. The independent variables are *Similarity_raw* and *Similarity_tfidf*. *D_Earnings Bad News* equals one if the *Earnings Bad News* in the five-day analyst report period is higher than the sample median in a given year and zero otherwise, where *Earnings Bad News* is calculated as the product of minus 1 and the distance between a firm's earnings announced in the five-day report window and the most recent analysts' I/B/E/S consensus earnings forecast, scaled by the security price two trading days prior to the earnings announcement. *D_Uncertainty* equals one if the *Uncertainty* of the analyst report is higher than the sample median in a given year and zero otherwise, where *Uncertainty* is calculated as the number of words in the analyst report that are classified as uncertainty words, scaled by the total number of words and aggregated to the firm-year level by taking the average. *D_Broker Size* equals one if the number of analysts in the broker is higher than the sample median in a given year and zero otherwise. *D_General Experience* equals one if the *General Experience* of the analyst is higher than the sample median in a given year and zero otherwise. Other explanatory variables include *Tone Rev, Forecast Rev, Recommendation Rev, Target Price Rev, Leader-Follower Ratio, Firm News, Prior |ABR|, |Earnings News|, Word Counts, Size, Market Value*, and *Book-Market Ratio*. For definitions of these variables, please refer to Table A-1. All continuous variables are winsorized at the 1^{st} and 99^{th} percentiles. For brevity, the coefficients on control variables are not reported. The firm, broker, and year fixed effects are added. The standard error is clustered at both the analyst and firm level. T-statistics are reported in parentheses. *, **, and *** indicate statistical significance at the 10%, 5%, and 1% levels, respectively.

	(1) |ABR|	(2) |ABR|
Similarity_raw	-0.226***	
	(-8.810)	
Similarity_raw*D_Earnings Bad News	-0.024***	
	(-2.757)	
Similarity_raw*D_Uncertainty	0.007	
	(0.877)	
Similarity_raw*D_Broker Size	0.019**	
	(2.062)	
Similarity_raw*D_General Experience	-0.006	
	(-0.643)	
Similarity_tfidf		-0.320***
		(-9.078)
Similarity_tfidf*D_Earnings Bad News		-0.028***
		(-2.726)
Similarity_tfidf*D_Uncertainty		0.007
		(0.811)
Similarity_tfidf*D_Broker Size		0.023**
		(2.083)

Similarity_tfidf*D_General Experience		-0.007
		(-0.633)
D_Earnings Bad News	0.012***	0.008***
	(3.291)	(3.500)
D_Uncertainty	-0.001	-0.000
	(-0.359)	(-0.022)
D_Broker Size	-0.006*	-0.004
	(-1.660)	(-1.433)
D_General Experience	0.003	0.002
	(0.891)	(1.071)
Controls	Yes	Yes
Firm fixed effects	Yes	Yes
Broker fixed effects	Yes	Yes
Year fixed effects	Yes	Yes
Number of observations	124515	124515
Adjusted R-squared	0.246	0.246

1.6.3. Interaction with corporate uncertainty

Previous research indicates that the analysts' research is more demanded for firms with a higher information uncertainty (Lobo et al., 2012; Peterson et al., 2015; Ye and Yu, 2017). Hence, investors might put higher weight on analysts' efforts to provide useful information to help them understand the firm. In this section, I examine whether investors' reaction to textual similarity among analyst reports is affected when the uncertainty of firms' information environment is higher. To do so, I use the word list of uncertainty developed by Loughran and Mcdonald (2011) to develop a measure of company uncertainty based on the analyst report transcripts. Specifically, I calculate the number of words in the uncertainty word list in each analyst report, scaled by the total number of words in the report. This method is also used by prior study to capture the company uncertainty reflected in the conference call transcripts (Huang et al., 2018). After obtaining the uncertainty score of each analyst report transcript, this score is aggregated to the firm-year level by taking the average (*Uncertainty*)

The indicator variable, *D_Uncertainty*, is defined as equal to one if *Uncertainty* is higher than the sample median in a given year and zero otherwise. I expand Equation (2) by including the interaction terms between each of the two similarity measures (*Similarity_raw* and *Similarity_tfidf*) and the *D_Uncertainty*. The results are reported in Table 1-8 below. I observe that the coefficients on the interaction terms (0.007 and 0.007) are not significant, suggesting that there is no difference in the investor reaction to analyst report similarity when the uncertainty of firms' information environment is greater.

1.6.4. Interaction with analyst characteristics

I study two following analyst characteristics to examine their impact on the association between analyst report similarity and the investor reaction: the size of brokerage house that employ analysts (*Broker Size*) and analysts' experience (*General Experience*). Based on prior literature (Stickel, 1995; Clement, 1999; Clement and Tse, 2005; Cao et al., 2023), analysts from larger brokers may have better data resources and better access to firm management, leading to more accurate forecasts. Thus, analysts from larger brokers have less concern about their work and are less likely to have herding behavior.

In addition, analysts with higher experience are less likely to be punished through termination for relatively poor forecast accuracy and for relatively bolder forecasts than their less experienced counterparts (Hong et al., 2000; Clement and Tse, 2005). However, literature also finds that younger analysts may have stronger incentives to deviate from the consensus to compensate for their deficiency in experience (Soltes, 2014; Frijns and Huynh, 2018). Therefore, it is not clear how analyst experience may affect analysts' herding behavior.

Based on the previous empirical findings, analysts from bigger brokerage houses or with higher experience in their work are less likely to take herding behavior. In addition, they are likely to possess better ability of collecting and analyzing corporate information. Thus, according to the hypothesis H2b, I predict that the observed negative association between analyst report similarity and the investor reaction is mitigated for these analysts. That is, even if these analysts provide textual information that is similar to prior analysts, such similarity is less likely to be caused by analysts' herding behavior or the lack of ability.

Accordingly, I introduce an indicator variable, *D_Broker Size*, which is equal to one if the *Broker Size* is higher than the sample median in a given year and zero otherwise. Another indicator variable, *D_General Experience*, is equal to one if the *General Experience* is higher than the sample median in a given year and zero otherwise. I expand Equation (2) by including the interaction terms between the two similarity measures (*Similarity_raw* and *Similarity_tfidf*) and measures of two analyst characteristics (*D_Broker Size* and *D_General Experience*).

Table 1-8 presents the results. The coefficients on interaction terms between the two measures of analyst report similarity and broker size (*D_Broker Size*) (0.019 and 0.023) are positive and significant at the 5% level in Column 1 and 2, suggesting that the investors' negative reaction to the similar textual information in an analyst report is moderated if the analysts are from bigger brokers. However, the coefficients on the interaction terms between the two measures of analyst report similarity and analyst experience (*D_ General Experience*) (-0.006 and -0.007) are not found to be statistically significant.

1.7. Robustness tests

1.7.1. Using abnormal trading volume

The abnormal stock return is used in the prior sections to capture the investor reaction. In this section, I use an alternative approach to measure it. Specifically, instead of the abnormal stock return, I measure the investor reaction by calculating the abnormal trading volume (*Volume*), defined as the natural logarithm of the cumulative trading volume over the three-day report window starting from the current report date, minus the natural logarithm of the firm-specific median trading volume for continuous three-day periods over the 365 days prior to the report date (De Franco et al., 2015).

Next, I estimate the regression model in Equation (2) by replacing the dependent variable (|*ABR*|) with *Volume*. The results are reported in Table 1-9. The coefficients on *Similarity_raw* and *Similarity_tfidf* (-2.149 and -3.020) are significantly positive at the 1% level in both Column 1 and 2. This observed negative relationship between the two measures of analyst report similarity and abnormal trading volume suggests that such textual similarity among analyst reports reduces the informativeness of analyst reports, leading to weaker stock reaction. Overall, these results are consistent with the baseline results in Table 1-5.

Table 1-9 Alternative Measure of Investor Reactions

This table presents the results of regressions of investor reactions on analysts' tendency to provide similar textual information in their reports. The dependent variable is *Volume*, defined as the natural logarithm of the cumulative three-day trading volume starting from the current report date, minus the natural logarithm of the firm-specific median trading volume for continuous three-day periods over the 365 days prior to the report date. The independent variables are *Similarity_raw* and *Similarity_tfidf*. Other explanatory variables include *Tone Rev, Forecast Rev, Recommendation Rev, Target Price Rev, Leader-Follower Ratio, Firm News, Prior |ABR|, |Earnings News|, Word Counts, Size, Market Value*, and *Book-Market Ratio*. For definitions of these variables, please refer to Table A-1. All continuous variables are winsorized at the 1st and 99th percentiles. For brevity, the coefficients on control variables are not reported. The firm, broker, and year fixed effects are added. The standard error is clustered at both the analyst and firm level. T-statistics are reported in parentheses. *, **, and *** indicate statistical significance at the 10%, 5%, and 1% levels, respectively.

	(1) Volume	(2) Volume
Similarity_raw	-2.149***	
	(-7.321)	
Similarity_tfidf		-3.020***
		(-7.373)
Controls	Yes	Yes
Firm fixed effects	Yes	Yes
Broker fixed effects	Yes	Yes
Year fixed effects	Yes	Yes
Number of observations	124515	124515
Adjusted R-squared	0.337	0.337

1.8. Concluding remarks

This research investigates analysts' tendency to provide similar textual information in their reports. The results provide empirical evidences that analysts are more likely to discuss similar textual information in their reports when they tend to have herding behavior, when they are less capable of collecting and providing new corporate information, when they are likely to learn from other analysts, and when there is significant firm-related news. Furthermore, the results show that an analyst report with more similar textual content to prior reports tend to be less informative, as reflected by the lower absolute value of three-day abnormal stock return immediately after the analyst report. Since this is a short-window event study, the observed impact is less likely to be caused by omitted variables, which is more likely to indicate a causal relationship.

The further results show that if an analyst report provides more similar textual information to other prior reports, the analyst's quantitative outputs (i.e., the earnings forecasts, stock recommendations, and price target forecasts) become less informative for investors. This is mostly consistent with that analysts' textual information in reports is considered as important by investors to understand the analysts' quantitative outputs, and that these quantitative outputs become less informative without new textual information in analyst reports to justify them.

I conduct several further tests and examine how this negative impact of analyst report similarity on the investor reactions varies across firm characteristics and analyst characteristics. I find that the negative relationship between analyst report similarity and the investor reactions is more pronounced when firm managers are more likely to withhold relevant information, but this negative relationship is moderated if analysts are from bigger brokers. In the robustness test, I measure the abnormal trading volume instead of abnormal return to capture the investor reactions. The results remain robust.

Overall, by shifting the focus from quantitative outputs to the textual content of analyst reports, this study provides new insights into analysts' tendency to exhibit similar behavior. Based on a large sample of analyst reports, this research shows that analysts' tendency of acting similarly extends beyond quantitative metrics to textual communication, which could be driven by several underlying mechanisms. The study further demonstrates that textual similarity among reports reduces their informativeness, emphasizing the importance of originality and uniqueness in textual content and providing new insights into how analysts' textual communication influences investor decision-making. The study also suggests that market investors value analysts' ability to discover new information not only beyond what firms disclose but also beyond what other analysts report. This research extends the understanding of analysts' information intermediary role, highlighting the importance of textual analysis in capturing analysts' behavior and its consequences.

One limitation of this study is related to the sample selection. The study is limited to S&P 500 companies from 2015 to 2020 due to the time constraints and difficulty for downloading extensive reports. This might restrict the generalizability of the findings to smaller firms or other time periods. Additionally, this study is conducted in the context of U.S. markets, which may have unique characteristics that do not generalize to other regions. Future research could examine the textual similarity and its impact on informativeness in different contexts. Finally, although the study employs a short-window event study method to establish causality, there may still be unobserved factors influencing both textual similarity and investor reactions.

Future research could concentrate on other textual attributes based on various textual analysis method, which can provide a more comprehensive understanding of how textual content influences investor decision-making. With the increasing use of artificial intelligence and machine learning in financial analysis, future research could explore how these technologies influence the textual content of analyst reports and their informativeness. In addition, future studies could expand the sample to include a broader range of firms, time periods, and regions to validate the results. Moreover, future research could use alternative methodologies, such as natural experiments or instrumental variables, to further address endogeneity issues.

1.9. Appendices

Table A-1 Variable Definitions

Variable	Description
Similarity_raw	The mean value of cosine similarity scores between the word frequency vector of an analyst report and those from the most recent reports issued by other analysts in the prior 60 days.
Similarity_tfidf	The mean value of cosine similarity scores between the word frequency vector of an analyst report and those from the most recent reports issued by other analysts in the prior 60 days, where the word frequency vector is weighted adjusted using term frequency–inverse document frequency (tf-idf) weighting method.
Word Counts	The natural logarithm of one plus the total number of words in an analyst report.
Negative	The number of words in the analyst report that are classified as negative using the negative word list from Loughran and Mcdonald (2011), scaled by the total number of words.
Positive	The number of words in the analyst report that are classified as positive using the positive word list from Loughran and Mcdonald (2011), scaled by the total number of words.
Determinants	
Leader-Follower Ratio	The product of minus 1 and the cumulative number of days by which the preceding two reports lead the report of interest divided by the cumulative number of days by which the subsequent two reports follow the report of interest.
Herding_forecast	The product of minus 1 and the absolute distance of an analyst's one year ahead earnings forecast for the firm from the average value of one year ahead forecasts in the prior 60 days issued by other analysts following the same firm, with this difference scaled by the security price two trading days before the forecast date.
Similarity_self	The cosine similarity score from the word frequency vector of an analyst report compared to the most recent report from the same analyst, where the word frequency vector is weighted adjusted using term frequency–inverse document frequency (tf-idf) weighting method.
Lag AFE	The absolute forecast error of the last one year ahead earnings forecast issued by an analyst for the firm in the prior year, where the absolute forecast error is calculated as the absolute difference between an analyst's forecast for the firm and the firm's actual earnings per share in the given year, scaled by the security price two trading days before the forecast date.
Broker Size	The number of analysts employed by the brokerage, as captured by the number of unique analysts from the brokerage who have issued at least one earnings forecast for any firm in a given year.
Forecast Frequency	The number of forecasts made by an analyst for the firm in a given year.
Firm Experience	The number of years an analyst has issued earnings forecasts for the firm.
General Experience	The number of years an analyst has issued earnings forecasts for any firms in the I/B/E/S database.
Firms Followed	The number of companies followed by an analyst following the firm in a given year.

Industries Followed	The number of industries followed by an analyst following the firm in a given year.
Number of Reports	The number of other analysts' reports for the same firm issued 60 days prior to the current analyst's report date.
Consensus Accuracy	The product of minus 1 and the mean value of the absolute forecast errors for the earnings forecasts issued in the prior 60 days, where the absolute forecast error is calculated as the absolute difference between an analyst's forecast for the firm and the firm's actual earnings per share in the given year, scaled by the security price two trading days before the forecast date.
Firm News	The number of analyst reports issued for the firm clustered over the five-day time window around the current analyst report.
Prior |ABR|	The absolute value of the cumulative ten-day abnormal returns ending the day before the current report date, with abnormal returns calculated as the buy-and-hold return minus the buy-and-hold return on the value-weighted market index for the same period.
Other variables	
|ABR|	The absolute value of the cumulative three-day abnormal returns starting from the current report date, with abnormal returns calculated as the buy-and-hold return minus the buy-and-hold return on the value-weighted market index for the same period.
Volume	The natural logarithm of the cumulative three-day trading volume starting from the current report date, minus the natural logarithm of the firm-specific median trading volume for continuous three-day periods over the 365 days prior to the report date.
Tone Rev	The absolute difference between the sentiment score of an analyst's report and the sentiment score of the prior report by the same analyst for the same firm, with the sentiment score calculated as the number of positive words minus the number of negative words, scaled by the total number of positive and negative words.
Forecast Rev	The absolute difference between an analyst's earnings forecast issued in the five-day report period and the analyst's prior forecast, scaled by the security price two trading days before the forecast date.
Recommendation Rev	The absolute difference between an analyst's stock recommendation issued in the five-day report period and the analyst's prior recommendation, where the analyst recommendation is assigned with numeric value by I/B/E/S database that a higher value represents a more negative recommendation (i.e., 1: Strong Buy; 2: Buy; 3: Hold; 4: Underperform; 5: Sell).
Target Price Rev	The absolute difference between an analyst's price target forecast issued in the five-day report period and the analyst's prior price target forecast, scaled by the security price two trading days before the forecast date.
|Earnings News|	The absolute value of a firm's earnings announced in the five-day report window minus the most recent analysts' I/B/E/S consensus earnings forecast, scaled by the security price two trading days prior to the earnings announcement, which is set to 0 if the firm does not make any earnings announcement in the five-day report window.
Size	The natural logarithm value of the firm's total sales in a given year.

Market Value	The natural logarithm value of the firm's market value at the beginning of the year.
Book-Market ratio	The book value of corporate equity divided by the market value of corporate equity at the beginning of the fiscal year.
Earnings Bad News	The product of minus 1 and the difference between the firm's earnings announced in the five-day report window and the most recent analysts' I/B/E/S consensus earnings forecast, scaled by the security price two trading days prior to the earnings announcement, which is set to 0 if the firm does not make any earnings announcement in the five-day report window.
Uncertainty	The number of words in the analyst report that are classified as uncertainty words using the uncertainty word list from Loughran and Mcdonald (2011), scaled by the total number of words and aggregated to the firm-year level by taking the average.

Table A-2 Investor Reactions to Analyst Report Similarity

This table presents the results of regressions of investor reactions on analysts' tendency to provide similar textual information in their reports. The dependent variable is |*ABR*|, defined as the absolute value of the cumulative three-day abnormal returns starting from the current report date, with abnormal returns calculated as the buy-and-hold return minus the buy-and-hold return on the value-weighted market index for the same period. The independent variables are *Similarity_raw* and *Similarity_tfidf*. Other explanatory variables include *Tone Rev, Forecast Rev, Recommendation Rev, Target Price Rev,* |*Earnings News*|, *Leader-Follower Ratio, Herding_forecast, Similarity_self, Lag AFE, Broker Size, Forecast Frequency, Firm Experience, General Experience, Firms Followed, Industries Followed, Number of Reports, Consensus Accuracy, Firm News, Prior* |*ABR*|, *Word Counts, Size, Market Value*, and *Book-Market Ratio.* For definitions of these variables, please refer to Table A-1. All continuous variables are winsorized at the 1st and 99th percentiles. The firm, broker, and year fixed effects are added. The standard error is clustered at both the analyst and firm level. T-statistics are reported in parentheses. *, **, and *** indicate statistical significance at the 10%, 5%, and 1% levels, respectively.

	(1) |ABR|	(2) |ABR|
Similarity_raw	-0.015***	
	(-4.700)	
Similarity_tfidf		-0.024***
		(-5.893)
Tone Rev	0.003***	0.003***
	(6.000)	(6.032)
Forecast Rev	0.334***	0.333***
	(5.185)	(5.164)
Recommendation Rev	0.003***	0.003***
	(3.515)	(3.443)
Target Price Rev	0.059***	0.059***
	(14.191)	(14.184)
|Earnings News|	0.116	0.114
	(0.655)	(0.648)
Leader-Follower Ratio	-0.000***	-0.000***
	(-4.930)	(-4.855)
Herding_forecast	-0.371***	-0.370***
	(-9.159)	(-9.130)
Similarity_self	0.002***	0.002***
	(3.478)	(3.609)
Lag AFE	0.007	0.007
	(0.145)	(0.134)
Broker Size	-0.000	-0.000
	(-0.786)	(-0.785)
Forecast Frequency	0.000	0.000
	(1.451)	(1.410)

Firm Experience	0.000	0.000
	(0.743)	(0.804)
General Experience	-0.000	-0.000
	(-0.048)	(-0.045)
Firms Followed	-0.000	0.000
	(-0.063)	(0.012)
Industries Followed	0.000	0.000
	(0.867)	(0.845)
Number of Reports	-0.001***	-0.001***
	(-9.344)	(-9.361)
Consensus Accuracy	-0.089***	-0.089***
	(-2.737)	(-2.744)
Firm News	0.001***	0.001***
	(17.084)	(17.129)
Prior |ABR|	-0.015	-0.015
	(-1.572)	(-1.556)
Word Counts	-0.001***	-0.001***
	(-5.000)	(-4.078)
Size	-0.004*	-0.004*
	(-1.652)	(-1.658)
Market Value	-0.001	-0.001
	(-0.383)	(-0.395)
Book-Market Ratio	0.009***	0.010***
	(3.619)	(3.627)
Firm fixed effects	Yes	Yes
Broker fixed effects	Yes	Yes
Year fixed effects	Yes	Yes
Number of observations	124515	124515
Adjusted R-squared	0.256	0.256

Appendix A-3 Pre-processing of analyst reports

Before conducting the textual analysis and constructing my models, it is necessary to merge the downloaded 456,596 analyst report transcripts with the data from other databases. The downloaded identifier information from the Thomson One database contains four kind of information about each analyst report: the firm's ticker, the release date of the report, the name of the analyst that issue the report, and the name of the broker that employs the analyst. However, it is significantly challenging to match each analyst report with the analyst forecast data from the I/B/E/S database. Ideally, an accurate matching requires at least the above four kinds of identifying information (i.e., the firms' ticker, the release date of the analyst earnings forecast, the name of the analyst, the name of the broker) from the I/B/E/S database. But, rather than the real name of the analyst or broker, the I/B/E/S only provides a 5-digit code to represent each broker and a 6-digit code for each analyst, yet what these codes really stand for are not disclosed. Thus, I use the following way to recognize the names of the brokers and analysts that these codes represent.

In order to do so, I firstly match each analyst report with all analyst forecasts from the I/B/E/S database that are issued for the same firm and on the same date. Among the 357 brokers, 81 of them cannot be matched based on the firm and date, and thus cannot be identified. This could happen for some very small brokers because not every analyst report is accompanied with the earnings forecast on the same date. This should not have a big impact on the sample because these 81 brokers are very small brokers and only provide 227 analyst reports in this sample (i.e., 0.05% of the 456,596 reports). Next, I calculate the rate of successful matching between the broker name and possible broker codes from the I/B/E/S database. For example, if there are 12 possible matching broker codes for the broker name, that is, there are earnings forecasts from 12 brokers for the firm issued on the same dates as the reports from the focal broker for the firm, then there could be 12 possible ways of matching. For each pair, I calculate the ratio of sample that are successfully matched to each other. I may observe that 35% of the analyst reports from the focal broker (e.g., EVERCORE ISI) in the Thomson One database are successfully matched to the earnings forecasts from the broker that have the broker code of '03374' in the I/B/E/S database. Following this procedure, I also use a reverse matching in that I match each analyst forecasts from the I/B/E/S database with all analyst reports that are issued for the same firm and on the same date. For example, I may observe that 20% of the earnings forecasts from the broker with the code of ‘03374’ are successfully matched to the analyst reports by the broker of ‘EVERCORE ISI.’ I next sum them up as a total rate of matching, which is 55%. A higher rate indicates a more successful matching between the broker and the code.

I choose the most possible combinations that have relatively higher rate of successful matching and make sure that both two constituent rates of matching are relatively high and comparable. During the manual checking, the most possible matching combination (with the highest total rate of matching) for each broker almost always have a much higher rate than the second best matching option. I also require the number of the analyst reports from the broker and earnings forecasts assigned to the broker code in the sample are similar. This could remove the cases that have high matching rate because some brokers with very few reports might accidentally be matched to another broker with a great number of earnings forecasts. And to increase the accuracy of matching, I further check the most possible matching (i.e., with relatively high rate of matching) for each broker to see whether, according to this combination, each analyst from the broker is almost exclusively assigned with a sole analyst code that is associated with the code of broker. Finally, I successfully distinguish and match 161 of all 227 brokers to the broker codes in the I/B/E/S. Specifically, these 161 brokers account for 362,111 of 456,596 analyst reports in this sample.

Appendix A-4 Processing of analyst reports

This section outlines the procedure of processing the downloaded analyst reports. Based on my observation, the structure and content of analyst reports can be rather complex, with different writing conventions and formats across analysts and brokers. Hence, to process the transcripts of analyst reports takes a lot of time and efforts.

First, during the downloading, the Thomson One database allows multiple reports to be downloaded simultaneously, with a maximum of 50 reports per download session. To shorten the time for data collection, I download as many reports as possible at each time. Therefore, each of the downloaded PDF documents contains more than one analyst report. Before processing the reports, each PDF document is split into individual analyst reports, I explore a Python program, the PyPDF library, to automatically split the PDF document into different analyst reports.

Next, the irrelevant information is excluded from the report. Specifically, I remove the tables, graphs, charts, and blank pages, which are not readable texts and do not contain textual information. Moreover, the company description, the introduction of broker and analyst, the glossary, the appendix, and the analyst disclosure and disclaimer are excluded. Such information does not contain analyst opinions, which is not relevant to the following research and could distort the outcome of textual analysis. To exclude such information from the report, I use Python to extract the content (i.e., the bookmark in the PDF document) of each analyst report. The content of the analyst report illustrates the structure and content of analyst report, such as the subtitles of the summary, tables, and graphics. Such content of analyst reports makes it possible to recognize those with undesired content and exclude them. However, at least for some reports, there is no subtitle for the textual information in the report transcript, while the subtitles of unwanted parts (e.g., tables, graphics, and analyst disclosure and disclaimer) show up. Thus, to directly remove the pages with only unwanted subtitles seems a little bit coarse, which might also remove the useful textual information. As a result, during this process, I follow a relatively conservative standard to exclude the unwanted bookmarks, for the purpose of avoiding removing the useful textual information.

Specifically, I remove the page in an analyst report with a subtitle that represents analyst disclosure, appendix, or glossary only when this page is among the final 40% of all pages and when other subtitles in this page also represent unwanted parts (e.g., tables and graphics). Thus, the pages with useful textual information is less likely to be accidently excluded. In the section below, I use other method to extract the useful information and further exclude the above undesired content from the analyst report transcripts.

After removing the irrelevant information based on the content (i.e., bookmark) in analyst report transcripts, I use Python to convert these updated PDF documents into HTML format and extract the textual information. Such conversion automatically identifies the tables in PDF documents, and structures them in a tabular form in the HTML documents. Furthermore, these tables have the labels of “table” in the HTML documents. This could help accurately locate the tables in analyst reports and exclude only the tables instead of other useful information. The similar procedure is implemented for other unwanted parts that are not readable texts, including the graphs and charts. After removing the unwanted information, I extract the textual information from each of the HTML documents and put it into the TXT documents, because the txt format is a more appropriate format to process the textual content. In very rare cases, the output TXT documents or the original HTML documents contain no word at all. To obtain the possible textual information in these reports, I use the pdfplumber library in Python to directly convert the PDF documents into TXT documents. However, this method using the pdfplumber library is not as accurate as the former method based on HTML documents in that it sometimes misses some of the textual information in the PDF documents. And it cannot be used to correctly remove the unwanted information, such as the content in tables. Hence, it is only conducted here as a supplement for the former method.

I then develop a Python program to further process the analyst reports for future textual analysis. The previous studies that conduct textual analysis on the large sample of analyst report transcripts have implemented rather similar and standard method to process the transcripts (Huang et al., 2014; De Franco et al., 2015; Huang et al., 2018; Bellstam et al., 2020), such as removing the stop words and non-word figures. I mostly conduct the procedure in the following studies (Huang et al., 2014; Huang et al., 2018), who has provided a relatively more precise method to process the analyst reports and increase the accuracy of textual analysis. I also manually read the reports to change several details of the method to make it more suitable for the further textual analysis (i.e., the calculation of textual similarity). Specifically, I use the following steps to process the transcripts of analyst reports (the TXT documents) and extract the textual information.

First, all words are converted into lower case. This is because the computer-based textual analysis method requires the precise input that it would treat the same word in lower case and in upper case (e.g., Analysis and analysis) as two different words. Second, I remove the non-English characters, including numbers, punctuations (except the full stop at the end of each sentence) and special characters, since this research concentrates on the similarity between words across analyst reports. This step also helps exclude the tables remained in the report, since most of the tables are related to the corporate financial and accounting condition such as balance sheets and income statements, the content of which are mostly figures. The blank pages, if any, are also removed.

Third, I convert some professional financial phrases into individual words. For instance, I convert "gross profit" into "grossprofit," "gross margin" into "grossmargin," and "operating margin" into "operatingmargin." This is because such financial terminology is commonly used in analyst reports and the meanings of these phrases are different from the meanings of individual words. For example, the meaning of "balance sheet" is quite different from the meanings of "balance" and "sheet" when looking at these two words separately. Thus, this conversion could help retain their correct meanings by treating each of the financial phrases as a single word, instead of treating them as separate words. This could increase the accuracy when doing textual analysis.

Fourth, I further remove the tables and analyst disclosure sections from the analyst reports. It is possible that some of the TXT documents of analyst reports still contain tables that cannot be excluded from the prior method (i.e., by using the content in analyst reports or HTML tags). After excluding numbers, the remaining tables (e.g., firm's balance sheets) may still contain a column of accounting words. Thus, my method is to locate the columns of the table that have very few words in each row without the full stop at the end. Then these columns are excluded from the text. Similarly, I locate the title of analyst disclosure section and exclude such section from the analyst report if such section appears near the end of the analyst report.

Fifth, I exclude the company names and tickers to avoid regarding them as normal words in the model. Sixth, I use the nltk library in Python to conduct the lemmatization. Specifically, the plural nouns are transferred into singular forms, and verbs into their present tense. Following the previous studies above, I do not transform the words into their root format (i.e., stemming) because financial words with the same stem probably have different meanings, such as "growth" and "grow," "profitable" and "profit," and "accounting" and "account."

At last, the stop words are excluded from analyst reports. The stop words are very frequently used in the text but have little economic meanings when looking at these words solely without the neighboring words around it. Some of the typical stop words are "a," "also," "are," "by," "for," and "of." The word list of stop words is collected from Bill McDonald's website.[11] I also exclude the geographic words and the date words, using the word lists from Bill McDonald. By excluding these words, I can concentrate more on the content that is more likely to reflect analysts' opinions, knowledge, or private information about the firm.

[11] The word list of stop words is provided on: https://sraf.nd.edu.

2. Analyst Coverage and Corporate Culture

2.1. Abstract

This study investigates the impact of analysts on corporate governance in the context of corporate culture. I first examine the association between the analyst coverage and the score of the firm's culture. The baseline results indicate a negative relationship between the two variables, suggesting that the companies with higher level of analyst coverage tend to have a weaker corporate culture. Further results suggest that this negative association is more pronounced for the long-term oriented cultural values than other cultural values. To deal with the potential endogeneity issues, I first employ an instrumental variable method. The results of the two-stage least squares model support that the analyst coverage has a negative impact on the corporate culture. Furthermore, I conduct a quasi-natural experiment based on two exogenous shocks to analyst coverage. Consistently, the results of the Difference-in-Difference model suggest that the analyst coverage has a negative influence on the firm's culture. These results are mostly consistent with the pressure effect that analysts could impose short-term pressure on firms, increasing management myopia and leading to a weaker corporate culture. Other additional results show that this negative relationship is mitigated for firms that are covered by more experienced analysts, for firms that could beat analysts' earnings forecasts, and for firms with a better corporate governance. Overall, this study provides the first empirical evidence about how analysts' participation could influence the corporate culture.

2.2. Introduction

The corporate culture, which refers to shared values, beliefs, or norms within the company, has great importance in guiding firm members' behavior. According to a recent survey upon 1,348 US corporate executives, the majority of these firms' executives regard the corporate culture as one of the top three value-enhancing drivers to long-term firm value (Graham et al., 2022a). They believe that a strong corporate culture, which refers to a set of values and norms that are widely shared and strongly held within the company, could reduce firms' tendency to attend myopic and value-destroying activities. Similarly, other empirical findings from prior research suggest that corporate culture plays an important role in corporate governance, which could mitigate the agency conflict and improve the execution effort and coordination among firm employees (Van den Steen, 2010a). A strong corporate culture is also found to encourage long-term oriented corporate activities (Quinn, 2018), and to improve firms' stability during financial crisis (Fang et al., 2023).

Despite the potential advantage of constructing a strong corporate culture, Graham et al. (2022a) have found that only a small portion of corporate executives (less than 20%) think that the culture in their companies is where it is required to be. Additionally, more than half of these executives (around 69%) have blamed the corporate leaders' underinvestment in corporate culture. Correspondingly, other than the impact of corporate culture, it is very important to provide more insight into what could influence the corporate culture. However, the empirical evidence of the determinants of corporate culture is relatively deficient, especially for the large-sample analyses. Recently, based on the new measure of corporate culture obtained from textual analysis tools, some researchers have investigated the effect of internal governance policies on corporate culture, such as fiduciary duty of loyalty and shareholder litigation risk (Hu et al., 2022; Jiang et al., 2022). It is unclear whether an external agency, financial analysts, have a positive or negative impact on the construction of corporate culture.

As information intermediaries, analysts are responsible for collecting firm-related information from various sources and providing their analysis outcome to the market. Prior research has found that analysts could significantly affect corporate behaviors such as corporate innovation, firms' credit activities, and firms' emissions of toxic pollution (Guo et al., 2019; Hallman et al., 2022; Jing et al., 2022). Based on these motivations, I concentrate on financial analysts to provide the first empirical evidence (to the best of my knowledge) about the impact of analysts on the corporate culture. This research topic is important because it sheds new light on how analysts affect the corporate governance.

Based on the literature, analyst coverage can have two opposite impacts on corporate behaviors. On the one hand, analysts could reduce firms' level of information asymmetry and serve as external monitors, leading to less profitability for insider trades (Ellul and Panayides, 2018), lower level of stock price crash risk (Kim et al., 2019), decreasing emissions of toxic pollution (Jing et al., 2022). On the other hand, analysts might impose short-term pressure (e.g., through their earnings forecasts) on firms, which could reduce firms' innovation input (Guo et al., 2019) and firms' corporate social performance (Qian et al., 2019)

Based on analysts' different roles, the analyst coverage could have opposite effects on corporate culture. First, by reducing information asymmetry and serving as external monitors to firms, analysts create an image about firms' culture, which pushes firm leaders to adhere to the corporate cultural values and helps attract potential investors and employees that value the firms' culture. In addition, by reducing agency conflict between executives and investors, the analyst coverage might encourage firm leaders to act for firms' long-term values, such as investing more time and resources in corporate culture, which tends to produce long-term benefits instead of immediate profits. In contrast, analysts could impose short-term pressure on firm leaders, pressing them to generate instant profits to meet the earnings targets. Hence, firm leaders could be less willing to invest time or resources in corporate culture, because doing so could be difficult, risky, and time-consuming, and is less likely to produce immediate profits.

To test the two competing hypotheses, I use a large sample measure of corporate culture proposed by a recent paper (Li et al., 2021b). Their method uses textual analysis (i.e., word-embedding model) based on the Q&A sections in firms' conference call transcripts to construct the scores of five cultural values (i.e., innovation, integrity, quality, respect, and teamwork) which are the top five corporate cultural values discussed and advertised by S&P 500 firms (Guiso et al., 2015b). Compared with prior method using survey- or interview-based tools, this new method creates a much larger sample size (8,995 unique firms) and mitigates the self-serving problem, compared with other textual measures based on company websites or 10-K reports. Their further validations also provide confidence by finding that this measure of the five cultural values could significantly explain firms' corresponding activities.

Next, I examine the association between analyst coverage and the corporate culture. The baseline results suggest that firms with higher level of analyst coverage tend to have a lower score of corporate culture, which is consistent with the pressure hypothesis that analysts can put short-term pressure on firms, resulting in firm leaders' underinvestment in corporate culture. Further results show that this negative relationship is more pronounced for the long-term oriented cultural values (i.e., culture of innovation) that are more likely to encourage pursuing relatively long-term benefits rather than short-term earnings.

Although the above results are mostly consistent with analysts' pressure effect, the potential endogeneity issue may distort the baseline result. For example, other unobservable explanatory factors for corporate culture might be correlated with whether analysts cover a firm or not. Additionally, firms with a strong culture are likely to attract certain analysts. To deal with this potential issue, I use two identification methods to examine whether this negative relationship is causal.

The first method is to conduct the two-stage least squares (2SLS) instrumental variable method based on a commonly used instrumental variable, expected coverage, which is firstly proposed by Yu (2008). This instrumental variable is calculated based on the number of analysts employed by a broker across years, which is less likely to be correlated with any firm characteristics that may be determinants of corporate culture. The results from the two-stage least squares model suggest that this negative relationship is significant and causal.

The second identification method is to implement a quasi-natural experiment, using brokerage closures and mergers as the relatively exogenous shocks on analyst coverage. These two events could directly change the number of analysts covering a firm, but they are almost unrelated to firm characteristics since brokers closing their business or acquiring another broker mostly reflect their own business strategies (Hong and Kacperczyk, 2010; Kelly and Ljungqvist, 2012). The results from the Difference-in-Difference model further support the baseline results, indicating a causal and negative relationship.

Finally, I examine whether this negative relationship could vary with several analyst and firm characteristics. The results in this section suggest that the observed pressure effect of analyst coverage on corporate culture is alleviated for the following subsamples: firms covered by analysts with higher level of general and firm-specific experience, firms that are more likely to beat analysts' earnings forecasts, firms with better corporate governance as reflected by the higher competitive market.

This study contributes to the literature in several ways. First, this research advances the large literature about the impact of financial analysts. According to the previous literature, the financial analysts may influence not only the investors' investment decision, but also the covered firms' business activities. Specifically, on the one hand, as an important information intermediary, the financial analysts could reduce the information asymmetry and act as external monitors to the firm management, which could lead to the reduced earnings management (Yu, 2008), the decreased managerial expropriation of outside shareholders (Chen et al., 2015), and the improved employee welfare (Bradley et al., 2022). On the other hand, the financial analysts are also criticized by the literature because analysts may impose short-term pressure on firm managers, which increases the managerial myopia and leads to decreased innovation activities (He and Tian, 2013) and limited investment in the socially responsible activities (Qian et al., 2019). In this context, my research makes contributions to the above debate by investigating the impact of analyst coverage on the construction of the corporate culture. By showing that analyst coverage is associated with a weaker corporate culture, this research provides new empirical evidence on the unintended negative consequences of analyst activities.

Second, my empirical evidence further extends the study of He and Tian (2013). They find that firms covered by a larger number of analysts are associated with decreased corporate innovation as captured by less patents and patents of lower quality. My study complements their findings by concentrating on the corporate culture of innovation, which represents a set of values and norms among the firm members that encourage corporate innovation. Consistent with the decreased patenting activities, my research further shows that the culture of innovation within the company is weaker, which not only provides a possible explanation about the reduced innovation activities, but also illustrates the underlying cultural mechanism through which analysts' short-term pressure influence corporate activities. Taken together, these empirical findings provide a more comprehensive understanding of how the financial analysts influence the corporate innovation.

Third, this study adds to the broad literature about the corporate culture. The previous studies provide numerous empirical findings suggesting that constructing a strong corporate culture could be beneficial for corporate business, such as improving operational efficiency and firm value (Li et al., 2021b), increasing firm stability during financial crisis (Fang et al., 2023), and reducing agency conflict (Van den Steen, 2010a). In contrast, there is relatively less empirical evidence about what could influence the corporate culture. For instance, based on survey-based method, Graham et al. (2022a) find that the corporate culture could be weaker due to several external forces (e.g., impatient investors, inadequate governance structures, and lack of attention to culture within an industry) and internal shortcomings (e.g., trust among employees and coordination challenges). Additionally, Graham et al. (2022b) explore such survey-based evidence and find that a weak corporate culture is likely caused by management's underinvestment in culture due to time and resource constraints. In addition, two recent papers have found that the weakened fiduciary duty of loyalty and shareholder litigation rights intensify the agency conflicts within the firm and, thus, have negative impact on the corporate culture (Hu et al., 2022; Jiang et al., 2022). My study builds on these insights by introducing financial analysts as a significant external agency that influences the corporate culture. By conducting the large-sample analysis, I demonstrate that the short-term pressure imposed by financial analysts leads to a weaker corporate culture. This finding highlights the critical role of external factors in shaping and influencing the corporate culture.

2.3. Literature review and hypothesis

2.3.1. Concept of corporate culture

As for the organizational culture, there is no conclusive definition of it possibly because of its abstract and ambiguous features, despite its potential importance within a company. However, there are nonnegligible similarities in how researchers discuss organizational culture. In the last several decades, when describing the organizational culture, researchers mostly use terms such as shared values, ideologies, assumptions, understandings, beliefs, and expectations within the organization, which provide norms or guidance for individuals in the organization (Schwartz and Davis, 1981; Sathe, 1983; Trice, 1984; Deshpandé and Webster Jr, 1989; Deshpandé et al., 1993).

Based on this discussion, subsequent researchers further highlight the importance of organizational culture as a social control system based on a set of shared values and norms throughout the organization, which defines what is important and the corresponding attitude and behaviors for members (Kotter and Heskett, 1992; O'Reilly and Chatman, 1996). In particular, the values are the big principles organizational members try to live up to, and norms are the more specific standards that guide the members' behavior, which could reveal whether organization members truly follow these ideals during the course of their work. As an example, integrity could be a desirable cultural value for some companies to achieve, yet it is a cultural norm that employees are encouraged to report bad news about the company, which demonstrates integrity. Thus, culture serves as a social control system that guides organization members' attitudes and behaviors, and the cultural values and norms should be aligned with each other for the culture to be effective. Hence, a strong culture indicates a set of values and norms that are widely shared and strongly held within the organization (O'Reilly, 1989).

Though not necessarily the perfect definition, this definition of organizational culture (i.e., shared values and norms) has, to some extent, grasped the essence of prior researchers' discussion about culture. This definition is frequently used by the subsequent papers (Chatman et al., 1998; Lauver and Kristof-Brown, 2001; Kerr and Slocum Jr, 2005; Chatman et al., 2014; Fang et al., 2023).

Based on the above perspective, the organizational culture is something an organization has and could be compared with different organizations. In the research area of corporate culture, this definition of culture is also widely used (Chatman et al., 2014; Guiso et al., 2015b; Li et al., 2021b). That is, the corporate culture represents a set of shared values and norms within the company. Van den Steen (2010a) propose a similar but narrower description that a company's culture refers to the degree to which the firm members hold similar beliefs about how business should be conducted. These beliefs and values remain even after all the original members have left (Van den Steen, 2010b).

As a complement to this definition, Graham et al. (2022b) and Graham et al. (2022a) have surveyed 1,348 corporate executives in North America and interviewed 18 other prominent business leaders in depth. Based on their investigations, executives' definition of corporate culture is consistent with the above academic definition. According to the corporate executives' responses, a firm's culture was often linked to how decisions were made in the company and how the group dynamics were managed. In addition, the corporate culture is characterized by these executives as "a belief system," "a standard of behavior," "a coordination mechanism," "how employees interact with one another," "norms around how people treat people," and "the tone for what type of company this is," which do not seem to deviate much from the commonly used academic definition of corporate culture (i.e., shared values and norms).

Furthermore, the culture could greatly differ across firms. And a company may possess different cultures. Thus, the researchers make their efforts to identify and label a relevant and comprehensive set of cultural values within firms, which helps describe corporate culture in more detail. For example, Denison and Mishra (1995) have conducted case studies on five firms and identified four cultural traits: involvement, consistency, adaptability, and mission. This classification is designed based on two contrasts: the contrast between external orientation and internal integration, and the contrast between change and stability. As an example, the culture of adaptation is more likely to reflect external orientation and change, while the culture of consistency focuses on internal integration and stability. Apart from that, the Organizational Culture Inventory (OCI) designed by Cooke and Lafferty (1987) has identified 12 cultural values: humanistic-helpful, affiliative, approval, conventional, dependent, avoidant, oppositional, power, competitive, perfectionistic, and self-actualizing. They further group these 12 norms into 3 cultural styles (i.e., constructive, passive/defensive, and aggressive/defensive) depending on two dimensions: a concern for people versus a concern for tasks, and to fulfill high-order satisfaction needs versus to maintain low-order security needs. However, Chatman and O'Reilly (2016) criticize that some of these cultural values (e.g., defensive) might be a mix of many different cultural elements, making it hard to understand their unique cultural values.

Since some of the identified cultural values seem to have similar meanings or to be correlated with each other, some researchers have used factor analysis to cluster different cultural values into several groups that capture their core features. This method could help exclude the irrelevant cultural values and group together the highly interrelated ones. In this context, O'Reilly et al. (1991) build a set of 110 cultural values (e.g., risk taking, being rule oriented, and fairness) that are collected from the review of prior research (mostly survey-based) into organizational culture from the academic and practical perspectives. Then they hand identify and remove the ones that are redundant, irrelevant, difficult to understand, or hard to discriminate, which leads to 54 final cultural values (e.g., being innovative, being team oriented, and being careful). To build the profiles of the cultures of firms, they conduct surveys on around 300 employees (i.e., employed accountants) from eight of the largest U.S. public accounting firms, and on 730 middle-level managers employed by a government agency. Then they ask these individuals to sort the 54 cultural values depending on how characteristic each of these values was for the culture of the respondents' companies. After collecting the survey results, O'Reilly et al. (1991) use the principal components analysis upon the obtained culture profiles of companies based on the 54 cultural items to extract the most representative cultural features. As a result, they develop eight groups of cultural values that includes 33 of the 54 cultural items. And the cultural elements within each group tend to be highly correlated with each other. These eight final cultural values are labelled as innovation and risk taking, attention to detail, orientation toward outcomes or results, aggressiveness and competitiveness, supportiveness, emphasis on growth and rewards, a collaborative team orientation, and decisiveness.

More recently, Chatman et al. (2014) has updated this set of cultural values. Specifically, they first modify 16 of the 54 original values that are found to be redundant and not representative. Based on the updated 54 cultural items, they use a similar method to capture firms' cultural profiles by conducting the surveys on around 2,000 employees from 56 large publicly traded high-technology firms in the United States and 100 smaller private firms that are headquartered in Ireland. Similar to the prior research, they conduct the principal components analysis on the survey-based data and ultimately derive six cultural values which includes 34 of the 54 cultural items and explains 44 percent of the total variance. The six values are labelled as adaptability, integrity, collaborative, results oriented, customer oriented, and detail oriented.[12] The difference in the final cultural values between these two studies is because these two studies conduct surveys on two very different sets of companies. Although the new label names are mostly different from those in the study of O'Reilly et al. (1991), their meanings and the corresponding constituent cultural items between them are

[12] According to Graham et al. (2022a), the Adaptability is defined as willing to experiment, fast-moving, quick to take advantage of opportunities, taking initiative. The Collaboration is defined as team-oriented, supportive, not aggressive, low levels of conflict. The Community is defined as respectful of diversity, community, and the environment, inclusive, caring, and open. The Customer-orientation is defined as listening to customers, being brand driven, taking pride in service. The Detail-orientation is defined as paying attention to detail, being precise, emphasizing quality and safety, being analytical. The Integrity is defined as high ethical standards, being honest, transparent. The Results-orientation is defined as high expectations for performance, focus on achievement, competitive, demanding.

similar.

Recent studies have started to explore the textual analysis tools to identify different cultural values. One significant advantage of this method is that it enables researchers to conduct relatively large sample analysis on firms' culture-related texts and empirically identify the cultural values from the perspective of companies. Guiso et al. (2015b) find that S&P 500 firms mostly discuss their corporate culture on their websites. Thus, they collect all the relevant words used by S&P 500 companies to describe the corporate culture on their websites in a given time. Then they aggregate 50 most recurring cultural values into nine categories by assigning different cultural values into one category if they are highly associated with each other.[13] The nine categories are labelled as integrity, teamwork, innovation, respect, quality, safety, community, communication, and hard work, the meanings of which are very similar to the above values identified by prior literature (e.g., Chatman et al. (2014)). Among them, five categories of cultural values are mostly discussed (advertised by at least 50 percent of the S&P 500 companies): innovation, integrity, quality, respect, and teamwork.[14]

Overall, the research into corporate culture has made long-term efforts to clarify its general concept and to identify different specific cultural values across firms and within firms.

2.3.2. Effect of a strong corporate culture

In the long run, corporate culture is important. The interviewed executives by Graham et al. (2022b) and Graham et al. (2022a) mostly rank the corporate culture as among the top

[13] Taking cultural value, Integrity, as an example, researchers check all the words associated with it across all companies, select the most associated ones (e.g., Ethics), and group them into one category (i.e., Integrity).

[14] According to Guiso et al. (2015b), innovation culture refers to the words like Innovation, Creativity, Excellence, Improvement, Passion; quality culture refers to the words like Quality, Customer, Meet needs, Commitment, Make a difference, Dedication; integrity culture refers to the words like Integrity, Ethics, Accountability, Trust, Honesty, Responsibility; respect culture refers to the words like Respect, Diversity, Inclusion, Development, Talent, Employees, Dignity; and teamwork culture refers to the words like Teamwork, Collaboration, Cooperation.

three value-enhancing drivers to firm value. Furthermore, many responding CFOs consider the corporate culture as more important than other factors such as financial health, competitive advantage, or company strategy. These executives also indicate that an effective culture (i.e., the corporate culture that encourages the behaviors that could help achieve firm's strategies) could improve productivity by attracting desired employees and by increasing employees' commitment and self-discipline. One important reason of it is that employees will eventually have to make decisions that cannot be adequately regulated in advance (O'Reilly, 1989). Having a strong corporate culture (characterized by a set of values and norms that are widely shared and strongly held within the company) thus helps employees find common goals and drive them to work towards the same targets.

The management literature provides a large amount of evidence that a strong organizational culture contributes to performance (e.g., (Denison, 1984; Camerer and Vepsalainen, 1988; Chatman, 1991; Kim and Yu, 2004)). As for the corporate literature, the empirical evidence also shows that a strong corporate culture could improve the business outcomes in the long-term period. Guiso et al. (2015b) find that that corporate culture of integrity is positively correlated with several business outcomes (i.e., productivity, profitability, industrial relations, and attractiveness to job applicants). Several studies suggest that companies with a strong corporate culture experience higher motivation and faster coordination for firm employees (Van den Steen, 2010a), less short-termism of corporate executives (Quinn, 2018), higher firm stability during financial crisis (Fang et al., 2023), and better performance during COVID-19 (Li et al., 2021a). Graham et al. (2022a) find from their interviews that firm executives greatly believe corporate culture can contribute to the long-term firm value, while firms with an ineffective culture are more likely to take myopic actions, such as delaying valuable projects to achieve consensus earnings and to boost short-term stock price. Li et al. (2021b) find large sample evidence that the firm with a strong corporate culture is significantly associated with improved operational efficiency, greater corporate risk-taking, better executive compensation design, higher firm value, and decreased earnings management.

In contrast, a strong culture may also limit the freedom of thought within the firm and could restrict the firm's ability to attract people with different talents to join the company. Sørensen (2002) argue that a strong corporate culture has an impact on organizational learning in response to internal and external change, which might affect the flexibility within the firm. Their results support this by showing that in relatively stable environments, firms with a strong culture have more reliable (less variable) performance, while in volatile environments, the reliability benefits of a strong cultures disappear. Van den Steen (2010a) find that a strong culture (according to his definition, the less differences in beliefs among firm members about the best way of doing things) could discourage information collection and experimentation. This is because employees with similar beliefs are less required to collect more information to "convince" other employees who possibly agree with him, or to experiment different actions and learn about the payoffs of these actions. And therefore, the company with a strong culture is more likely to attract people with similar values or beliefs but discourage people with great talents yet different thoughts.

Furthermore, some researchers find that certain kinds of corporate cultures do have their side effects, which might cause damage to the firm. For instance, Balthazard et al. (2006) find significantly negative correlation between dysfunctional defensive corporate culture and the role clarity, communication quality, "fit," and job satisfaction within the firms, which implying a link between dysfunctional corporate culture and deficits in operating efficiency and effectiveness. Hutton et al. (2015) show that firms with a Republican culture are more likely to be sued for civil rights, labor, and environmental violations than Democratic firms, whose party ideology promotes equal rights, labor rights, and environmental protection. Liu (2016) find that a corporate culture which is more tolerant to corruption could encourage corporate misconduct including earnings management, accounting fraud, option backdating, and opportunistic insider trading.

Overall, the literature in this area provides much evidence that a strong corporate culture, however defined and measured, tends to produce long-term benefit and performance by guiding the way organizational members think and behave. Meanwhile, the firms have to be careful with their unique corporate culture and the possible bad effects.

2.3.3. Determinants of corporate culture

With respect to the determinants of corporate culture, there is not much empirical evidence about what could affect corporate culture, compared to the relatively extensive research into the impact of corporate culture. Despite that, the literature has provided some useful insight into this issue. First, many of the prior papers have assumed that firm leaders are almost the primary determinants of corporate culture, which is intuitively reasonable and to some extent taken for granted by the literature. Accordingly, a recent study by O'Reilly et al. (2014) conducts a survey on employees from 32 high-technology companies to obtain the profile of the corporate culture and the personalities of the firm CEOs. The responses suggest that the five CEO personalities (i.e., Extraversion, Agreeableness, Conscientiousness, Neuroticism, and Openness to Experience) could significantly influence the culture of a firm. For example, O'Reilly et al. (2014) find that the firm CEOs who are more open are associated with the corporate culture that is more characterized by adaptability. Their logic is that these personalities determine CEOs' course of behavior which consequently affects employees' perceptions about what matters and how to behave in the workplace. In addition, Davidson et al. (2015) find that the financial reporting risk increases during the tenure of unfrugal CEOs, and this increase is positively associated with executives' equity-based incentives, the presence of an unfrugal CFO, and a weak board monitoring. These empirical findings are consistent with their argument that an unfrugal CEO is associated with a culture characterized by a relatively weak control environment.

More recently, Graham et al. (2022b) implement a relatively extensive survey by asking questions upon 1,348 corporate executives. Consistent with the above studies, Graham et al. (2022b) also find that the current CEOs are regarded by the respondents as the most influential individuals in establishing the firm's current culture than other parties such as shareholders, founders, and non-management employees. Their evidence also suggests that the board members might play a role in establishing corporate culture, by taking direct actions like monitoring employees' daily actions, or by using indirect methods like selecting the CEO with desired characteristics.

Graham et al. (2022b) also ask about what could prevent the corporate culture from being strong. The highest ranked problem listed by these respondents is that the firm management might underinvest in the corporate culture. Consistent with the study of Schein (2017), such investments may include: (1) reaching consensus on firms' cultural goals; (2) developing standards to measure how well the goals are achieved; (3) having proper cultural reward and punishment systems (e.g., an appropriate compensation scheme) to regulate how employees behave during work and to attract the suitable people; (4) constructing group boundaries to define who belongs to the company and shares its culture; (5) developing acceptable norms around distributing power, authority, and influence within a company; and (6) disciplining peer relationship (i.e., trust and openness) among employees, especially for people of different cultures.

When asked about reasons for underinvestment, some surveyed executives suggest that investing in culture is risky and time-consuming, which is not consistent with their own interests. They rather prefer to do nothing about the corporate culture and only pursue low risks and immediate profits to avoid mistakes or failures and the significant punishments. These corporate executives also indicate other factors including whether other firms within the industry pay attention to culture, limitations on the time or resources, and whether there is an opportunity for change. In addition, some executives suggest that the impatient investors could lead to ineffective corporate culture. For instance, investors might wish the company to grow rapidly, regardless of whether growing too fast could damage the corporate culture. The result in their study also finds that a more concentrated ownership structure is negatively associated with corporate culture of integrity, which could suggest that the value-maximization of shareholders is not consistent with building a strong culture of integrity.

Furthermore, the literature suggests that the top executives themselves should live up to the cultural values to "lead by example" (e.g., by showing what they value and pay attention to) so as to share such culture in the community (Guiso et al., 2015b). For example, to construct the culture of innovation, top executives might probably want to "lead by example" by encouraging innovation activities within companies. This might include, but not limited to, increasing the Research & Development spending, hiring more R&D employees, acquiring other innovative firms, and investing in start-ups related to their core business. However, these activities tend to be risky and long-term, and are less likely to generate a short-term benefit (Hirshleifer et al., 2012; Moshirian et al., 2021). Thus, to build a strong corporate culture through "leading by example" might not always be aligned with executives' short-period interests (e.g., to meet near-term earnings targets and to increase stock price return), especially for the long-term oriented culture that represents the shared values and norms within the company to achieve long-term profits rather than short-term profits.

Some recent empirical results also provide evidence for executives' incentives that work against investing in corporate culture. Doukas and Zhang (2021) suggest that to invest in corporate social culture is not always consistent with firm managers' interests because such investments might not generate significant short-term profits. They find that low-ability managers are less willing to invest in corporate social culture than high-ability managers who pursue long-term strategies and long-term interests. In addition, Hu et al. (2022) show that the fiduciary duty of loyalty, which requires corporate fiduciaries (i.e., directors, officers, and dominant shareholders) to subordinate their own interests to their companies' is negatively associated with the corporate culture, suggesting that such duty intensifies agency conflicts and potentially erodes corporate culture. Moreover, Jiang et al. (2022) find that weaker shareholder litigation rights is associated with a weaker corporate culture, consistent with their argument that the weaker shareholder litigation rights leave directors and executives to engage in self-serving actions, which weakens the corporate culture.

Apart from the leadership, other factors might play a role in shaping corporate culture. For example, Chatman and Jehn (1994) show that firms in industries with more complex technology and with faster growth are correlated with higher levels of the culture of innovation and team orientation, and lower level of the culture of stability. Additionally, Guiso et al. (2015b) find that the public firms tend to have lower level of integrity culture than the private firms, suggesting that it could be beneficial for public firms to invest less in their culture of integrity, at least in the short-term period. Using survey-based method, Graham et al. (2022a) and Graham et al. (2022b) ask surveyed executives to rank the factors that determine corporate culture. The corporate marketplace is ranked as the second most important determinant (compared to the leadership). When the market environment changes (e.g., getting more competitive), it could require a different culture instead of insisting on the past way. Other possible determinants of the appropriate choice of corporate culture include the lifecycle of the firm, corporate industry position, and demographics of the workforce. In particular, a young and growing corporate is likely to pursue a more passionate culture, while a big and mature firm might select a relatively stable culture. And a company as market leader could probably prefer a more defensive culture. The appropriate corporate culture also depends on the demographics of employees who might differ in, for example, habits and preferences.

The respondents from this survey also discuss relevant issues that could lead to ineffective culture, which include: (1) culture does not follow the business changes and business requirements; (2) firm policies are inconsistent with the desired culture; and (3) employees have inefficient interactions with each other and less commitment to the corporate culture. In general, there is no universal cultural value that could appropriately apply to all firms and even within the same firm at all periods of time. Therefore, it is not an easy task to determine the appropriate culture for a firm and to strengthen it.

Overall, the literature provides intriguing insight into the role that firm leadership and other factors play in constructing the corporate culture. For firm leaders, to construct a strong corporate culture are likely to be difficult, risky, and time-consuming, which requires continuous investments but may not generate benefits in relatively short-term period.

2.3.4. Approaches for measuring corporate culture

The corporate culture is intangible, making it difficult to construct the appropriate and universally accepted measure to capture it. Prior researchers have tried several ways to measure it. One relatively direct way to capture the corporate culture is to implement tools like surveys, questionnaires, and interviews upon corporate members.

Jung et al. (2009) have conducted a review study and identified 70 ways to measure corporate culture used by prior researchers and most of these methods are based on surveys, questionnaires, or interviews. To understand the cultural values, many of them have further defined different sets of cultural dimensions to focus on specific cultural values of interest. For example, as discussed previously, a frequently used approach called Organizational Culture Profile (OCP) is introduced by O'Reilly et al. (1991). Their method is to collect the assessments of the corporate culture from the respondents that work in the companies. These respondents are asked to sort the 54 identified cultural items that are most characteristic or uncharacteristic of their organizations. These 54 cultural items are finally grouped into eight final cultural values. As a result, O'Reilly et al. (1991) could obtain the culture profile (i.e., OCP) of each company in this survey. This method is updated by Chatman et al. (2014), who modify 16 of the 54 cultural values that are regarded as redundant or not typical and group these cultural items into six final cultural values. Chatman et al. (2014) conduct further tests to validate their measure of the culture of adaptability. Their results show that this measure of the corporate culture of adaptability is positively associated with adaptability-related words in the 10-K reports and with the change in the research and development investments.

Graham et al. (2022a) collect executives' responses about their descriptions of the current culture of the firms for these executives, which are hand-coded into seven elements of cultural values (i.e., adaptability, integrity, collaborative, results oriented, customer oriented, detail oriented, and community-orientation). Similar to the above measures of cultural values in prior studies, Graham et al. (2022a) also measure to what extent employees have followed the corresponding cultural norms by asking direct survey questions about employees' actual workplace behaviors.

Some researchers have used the survey data collected from other available sources. One of the commonly used survey data is from the "100 Best Companies to Work For in America." During this survey, firm employees are anonymously asked a list of questions, such as how they feel about the firm management and whether they are satisfied with their jobs. The list of firm rankings is based on the employees' responses. In this context, Edmans (2011) has used the companies in this list to capture the companies with higher employee satisfaction. The results of his study show a positive relationship between this survey-based measure of employee satisfaction and the long-term stock return, suggesting that the market is not efficient enough to account for this intangible asset. Similarly, Bargeron et al. (2015) use this survey data to build their proxy for the corporate culture of trust, that is, the trust of employees in firm management. They find that the relative size of acquisitions announced by these firms with stronger trust culture is smaller than that announced by other firms, which is consistent with their argument that the firm's culture of trust is regarded as a valuable asset and that the firm with stronger trust culture tend to have M&A policies that are more likely to help protect this asset..

Taken together, it is easily noted that the advantage of using a survey-based method is that the responding employees or executives could directly express their views of corporate culture. However, this method is subject to some concerns. For example, this method could have self-selection problem because not every contacted informant (e.g., executives or employees) is happy to respond to the survey. The willing respondents might be from firms that are more interested in corporate culture or the firms that tend to have a better corporate culture. Moreover, the responses could be respondents' self-serving stories instead of revealing their actual thoughts.

More importantly, the nonnegligible concern of this survey-based method is that it severely limits the sample size for research, possibly because the information collection is time-consuming and only feasible for a small number of willing respondents that work or might not work in the concerning companies. For example, Chatman et al. (2014) only collect corporate culture assessments from respondents in 156 companies. Graham et al. (2022a) have sent the survey requests to a larger list of firm executives' email addresses provided by Fuqua School of Business at Duke University and Columbia Business School. But the response rate of their survey is only 13.4%, which is already a relatively higher rate compared to prior research (Graham and Harvey, 2001). And these responding firms are concentrated on the firms with larger size, higher leverage, greater profitability, lower sales growth, and higher credit ratings compared with the Compustat firms. Thus, the associations observed from this method could be caused by these characteristics or other omitted features, making it questionable to apply these findings to the more generalized sample. Considering this potential issue, it is very important to construct the new measures of corporate culture that could cover a much large sample of companies, including the firms that tend to have a weak culture and the firms that do not care about their corporate culture.

Several studies have started to use other public data to construct the large sample measures of corporate culture. Their methods are used to capture various dimensions of corporate culture. For example, Hoi et al. (2013) suggests that corporate social responsibility (CSR) could be viewed as a facet of corporate culture or, specifically, the culture of taking the "right" course of action that considers not only the economic, but also the social, environmental, and other externalized impacts of the company's activities. Their findings suggest that the companies with a weaker CSR-related culture (as reflected by higher level of irresponsible CSR activities) are more likely to take aggressive tax avoidance activities. In addition, as discussed previously, Davidson et al. (2015) use the CEOs' past unethical behavior to proxy for the corporate culture that has weak discipline and controls, which is found to be associated with higher financial reporting risk. Hutton et al. (2015) measure corporate political culture (preferring Republican Party or Democratic Party) by calculating the difference between the political contributions to either parties. Their results indicate a higher likelihood of being sued for civil rights, labor, and environmental violations for companies with a stronger Republican culture.

These new proxies for corporate culture have a significant advantage of having a much bigger sample size. However, the accuracy of them might be limited because these variables are likely to capture something more than corporate culture. In particular, the measures based on CSR, CEOs' past experiences or political preferences might be a bit general to capture something else. Even if these measures could capture corporate culture, they possibly contain more than one cultural value. In addition, each of these measures is used to only capture the certain kind of corporate culture, so it is not feasible to use similar methods to capture other important corporate cultures or to compare the measures of various corporate cultures. Meanwhile, the above methods suggest there is a need to conduct validation tests to determine whether the proposed measures could accurately capture some cultural values.

Recently, the advances in textual analysis method introduce new ways to develop the large sample measures of corporate culture while mitigating some of the above problems. A recent study by Graham et al. (2022b) recommends using textual analysis method on the firm-related texts (e.g., conference call transcripts, analyst reports, and company's external communication). For example, the researcher may use the company's call transcripts to pick out the corporate executives' words, what they discuss, what they value, and how they talk about other employees. But the researchers have to determine which words can best describe various corporate cultural values.

With respect to the textual analysis upon firm texts, Guiso et al. (2015b) have investigated the websites of S&P 500 companies and found that these firms mostly talk about their corporate culture by identifying a set of central cultural values (e.g., integrity) and several key words to explain them (e.g., honesty and transparency). Guiso et al. (2015b) download the textual content that describes corporate culture from the website of each S&P 500 company. They next conduct the textual analysis on these culture-related texts. Among all these cultural values, 50 most commonly observed values are chosen and are grouped into nine categories by aggregating different values into one group if they are highly correlated with each other. These nine advertised cultural values are labelled as integrity, teamwork, innovation, respect, quality, safety, community, communication, and hard work.

This process reduces the large set of cultural values to only several groups that are more representative and meaningful. Guiso et al. (2015b) define the dummy variable for each of these nine advertised cultural values depending on whether a company advertises such cultural value on its website. After that, they examine the relationship between each cultural value and the firm's performance. However, they couldn't find significant evidence that these dummy variables of nine advertised cultural values are associated with company performance, possibly because the words on company websites are more likely to reflect the firm's advertisement of their culture rather than the actual profile of corporate culture.

In addition, Audi et al. (2015) conduct the textual analysis on the Management Discussion and Analysis (MD&A) sections in corporate 10-K reports to capture the culture of trust within a company. They hand select 21 "trust" related words (e.g., trust, integrity, ethics, and honest) and calculate the number of times these words appear in MD&A sections of 10-K reports as the proxy for the corporate culture that involves trust. Their sample includes a large number of public firms, which leads to a sample of 40,441 firm-year observations for 15 years. Audi et al. (2015) next investigate the implication of companies' trust culture for the stock market and find that the firms with higher score of trust culture tend to have a higher level of stock price volatility. According to their theory, this is because firms with a trust culture could have larger downside risk in that it is difficult to recover the trust once it is broken.

Similarly, a more recent study by Fang et al. (2023) focuses on the texts in corporate 10-K reports. To calculate the measures of a firm's collaborative, controlling, competitive, and creative culture, they select a set of synonyms for each of the four cultural values from the prior literature and the Harvard IV-4 Psychosocial Dictionary. For instance, words like "teamwork" and "cooperate" is correlated with the collaborative culture. After selecting the word list for the four cultural values, they calculate their measure of each cultural value as the frequency of words in 10-K reports that are related to each cultural value, scaled by the total number words in the reports. In addition, as a validation test for this measure, they find significant correlation between these four corporate cultural values and the firm characteristics related to employee relations, risk management, growth, and innovation. In the additional analysis, Fang et al. (2023) find that firms' controlling culture is positively associated with firm stability, as reflected by better performance (e.g., less decrease in asset, higher debt issuance, and better access to credit) during the financial crisis.

These above methods make it feasible to build the measures of corporate culture that could apply to a significantly larger sample of companies than the survey-based methods. More importantly, these measures based on the textual analysis approach could capture the more specific cultural values of firms (e.g., the culture of teamwork), compared to the prior indicators (e.g., CSR scores) that possibly capture something more than corporate culture. However, these new measures that use the textual analysis upon the corporate websites or the firms' 10-K reports are criticized in that the words on such websites or reports are likely to reflect executives' cheap talks to advertise their desired culture and to attract potential investors. Thus, the texts from these two sources might not accurately represent the true level of a firm's culture. In this context, the following research conducts the textual analysis based on a different kind of firm-related documents, which could, to some extent, mitigate the above concern.

Specifically, Li et al. (2021b) employ the textual analysis method (i.e., word embedding model) upon the Q&A sections of public firms' conference call transcripts to calculate the scores of five cultural values (innovation, integrity, quality, respect, and teamwork), which are the most advertised cultural values by S&P 500 firms (by more than 50 percent of these firms) on the culture sections of their websites. The choice of the companies' call transcripts is based on the argument that if corporate executives live up to the advertised cultural values and "walk the talk," their conversation during the calls are expected to reflect such values. Additionally, as indicated by the survey-based findings (Graham et al., 2022b), the firm leaders (e.g., CEOs) are regarded as the most influential person in constructing a firm's culture. Hence, the words used by firm leaders during the conference calls are expected to reflect the prevalent cultural values within their firms. Moreover, this measure uses the Q&A sections of public firms' conference call transcripts instead of the presentation sections because, as suggested in the prior literature, the Q&A sections are more informative than the presentation sections in that the active analyst involvement increases the information content of the calls (Matsumoto et al., 2011).

This method has several advantages that could mitigate some of the above issues in the previous measures. First, unlike the corporate websites, where companies could easily advertise or even boast about their corporate culture (e.g., intentionally exaggerating their corporate culture), the earnings conference call is not intended to talk about a firm's culture but the details of its business and performance, which could provide an opportunity to observe what firms' executives actually do than what they simply talk, thus mitigating the self-serving problem related to the advertisement of corporate culture. And compared to the textual content in 10-K reports, the conversation in the Q&A sections during firm calls are more spontaneous, which makes executives less likely to choose discussion topics in advance or to conduct window dressing (Li et al., 2021b).[15] Moreover, this calculated score of corporate culture is further validated and examined. The details of the calculation of this culture score, its validation tests, and robustness checks are discussed in the subsequent sections.

2.3.5. The role of analysts in corporate culture

As information intermediary, financial analysts collect information about the firm, analyze the future performance, and provide their analyst reports to the market participants. They provide value to capital market by confirming and interpreting public information, and discovering new information that is not easily available for market investors (Chen et al., 2010; Livnat and Zhang, 2012; Huang et al., 2018). One way of collecting the corporate information for analysts is to directly talk with corporate executives during earnings calls, where they are able to ask questions about corporate business operation and performance.

[15] The responses of firm executives to questions during the Q&A sections might not be absolutely spontaneous. Some of the prior studies show that firm management could make efforts to anticipate questions and prepare for them (Abraham and Bamber 2017; Bamber and Abraham, 2020).

Then analysts express their opinions about the covered firms to the clients through analysis outputs such as the research reports, earnings forecast, and stock recommendations. The literature suggests that these analyst research outputs, including the above three qualitative and quantitative outputs, are informative for market participants (Ivković and Jegadeesh, 2004; Asquith et al., 2005; Huang et al., 2014).

Similar to the above discussion, relatively extensive studies have examined analysts' role in security price formation. However, because of analysts' role as information intermediary and their interactions with the investors and companies, it is possible that analysts' participation could also have an impact on firms' behaviors. Although relatively not much literature has discussed the effects of financial analysts on corporate governance, the prior literature suggests that analysts could serve several different roles in corporate governance.

First, analysts could reduce firms' level of information asymmetry and serve as external monitors. Analysts represent the interests of not only corporate shareholders, but also potential investors and other market participants. They regularly meet and talk with corporate executives to collect information related to business. Their knowledge about accounting and finance, together with their industrial experience, enables them to understand the information from executives and corporate disclosures (e.g., the financial report). Then analysts provide their analysis outcomes for market participants to help them understand the corporate business and detect any misconducts. Thus, analysts reduce the information asymmetry between executives and market participants (e.g., corporate shareholders, potential investors, and other market participants), and serve as external monitors to executives. Accordingly, literature suggests that firms covered by more analysts tend to acquire more innovative firms and invest more in the corporate venture capital (Guo et al., 2019), to manage their earnings less (Yu, 2008), to have less managerial expropriation of outside shareholders (Chen et al., 2015), to decrease emissions of toxic pollution (Jing et al., 2022), and to improve employee welfare (Bradley et al., 2022).

In contrast to the above impact, the literature indicates that analysts could impose short-term pressure on firms. Specifically, analysts are responsible for predicting the near-term earnings and making corresponding stock recommendations. In addition, higher analyst coverage draws much attention of market participants to a firm and its corporate executives. Under such circumstance, if the firm fails to meet the analysts' short-term expectation of corporate earnings, this is found to be associated with penalties such as lower annual bonus for firm executives (Matsunaga and Park, 2001), bad influence on the reputation of firm executives (Graham et al., 2005), and negative stock market reaction (after controlling for firm's absolute performance) (Bartov et al., 2002; Kasznik and McNichols, 2002). The negative market reaction to the companies' failure to meet the target is found to be larger when firms have greater analyst coverage (Huang et al., 2017b). Therefore, firms with higher analyst coverage that are subject to more short-term pressure tend to disregard some long-term oriented projects which probably cannot produce short-term benefits. Consistently, it is found that the firms with higher analyst coverage are associated with investing less in the innovation project (Guo et al., 2019), producing less innovation output (He and Tian, 2013), and limiting investment in socially responsible activities (Qian et al., 2019), for the purpose of achieving short-term earnings targets.

With respect to the corporate culture, according to my knowledge, there is no empirical evidence about how analysts' participation might affect a firm's culture. According to the above discussion, higher analyst coverage can possibly have two opposite impacts on the corporate culture. On the one hand, analysts play a crucial role in reducing information asymmetry within firms, and acting as external monitors. By providing detailed insights into a firm's operations, strategies, and priorities, analysts enable market participants—such as investors, stakeholders, and potential employees—to gain a clearer understanding of the company's activities. Although analysts may not explicitly discuss corporate culture in their reports, the information they disclose can offer indirect clues about the values and norms held by the company. For instance, through highlighting the company's commitment to innovation, product quality, or employee welfare, analysts can help stakeholders infer the underlying cultural values that guide the organization.

This transparency may incentivize firms to align their practices with their publicly stated cultural values, rather than merely adverting them superficially, since firms know that their actions are being closely scrutinized by analysts and, by extension, the market. Moreover, reduced information asymmetry can attract potential stakeholders (e.g., investors and employees) who share the firm's cultural values. For example, investors are more likely to support companies whose values align with their own, while employees are more likely to thrive in environments where they feel cultural alignment. Such mutual reinforcement can further strengthen the corporate culture

Furthermore, a lower level of information asymmetry has been shown to mitigate agency conflicts between executives and investors (Irani and Oesch, 2013; Chen et al., 2015). When information flows more easily, executives are less likely to prioritize their own interests over those of the firm and its stakeholders. Analysts' involvement can encourage firm leaders to focus on long-term value-enhancing activities, even if these do not yield immediate profits. In such an environment, firm leaders may be more willing to invest time and resources in constructing a strong culture (e.g., setting cultural targets, establishing cultural reward and punishment systems, and monitoring the culture within the company), even though these investments may not produce instant benefits. Leaders who prioritize cultural development are also more likely to "lead by example," by living up to the firm's values in their actions and decisions, even when this does not yield short-term financial profits.

Taken together, the above arguments would suggest that firms covered by more analysts would have a stronger corporate culture (i.e., values and norms widely shared and strongly held by firm members). The first hypothesis in this study is given as below.

Hypothesis 1a. Analyst coverage is associated with a stronger corporate culture within the covered firms.

On the other hand, as discussed previously, analyst coverage can impose significant short-term pressure on firm leaders. The need to meet earnings targets may force leaders to prioritize immediate profits over long-term cultural investments. If they fail to achieve the short-term earnings targets, they will receive punishment such as reduced stock prices, decreased compensation, or even job risks. The resource constraints further exacerbate this issue. Firms often operate with limited resources, and leaders must make difficult trade-offs between short-term financial performance and long-term cultural development, given that these two elements are not always compatible with each other. Under the pressure from analysts, leaders may select the actions that generate immediate financial profits, even if these actions might not be optimal for the firm's long-term development.

In such an environment, leaders may be less inclined to invest in initiatives that strengthen corporate culture, such as defining the cultural goals, designing proper cultural reward and punishment systems, or monitoring the culture within the company. As discussed in Section 2.3.3, building a strong corporate culture is a complex, time-consuming, and risky process that requires sustained investment. Hence, when short-term earnings take precedence, leaders may neglect these efforts, leading to a weaker corporate culture.

Collectively, the firms with more short-term pressure are less likely to develop a strong culture. The second hypothesis in this study is given as below.

Hypothesis 1b. Analyst coverage is associated with a weaker corporate culture within the covered firms.

2.4. Sample selection

To investigate the association between analyst coverage and the corporate culture, I collect the data from several databases. The analyst forecasting data is collected from the Institutional Brokers' Estimate System Database (I/B/E/S). This database includes analyst level data and are mainly used by prior studies upon financial analysts. In particular, it provides analyst's earnings forecasts, recommendations, and target prices for each firm in a year, as well as the codes of analysts' names and brokers hiring them. The corporate accounting data are obtained from the COMPUSTAT Database, and stock market price data are collected from the Center for Research in Security Prices Database (CRSP). These two databases are primarily used by the prior literature to obtain the company data.

Moreover, I collect the corporate culture data from Li et al. (2021b).[16] They create a firm-year measure of the corporate culture by calculating the scores of five cultural values: innovation, integrity, quality, respect, and teamwork. To capture these five cultural values, they use a textual analysis tool to create culture-related word lists based on the texts in firms' conference call transcripts. Then they calculate each of the five culture scores based on the word list that is associated with each cultural value. The data of their scores of the corporate culture includes the US public firms from 2001 to 2021. Their method is discussed in detail in the next section.

The final sample only includes the firms that appear in all databases above. The firm-year data with missing value on any of the dependent variables, independent variable, and control variables are excluded. As a result, the final sample has 37,263 firm-year observations.

2.4.1. Measuring corporate culture

As discussed in the Section 2.3.4, the abstract nature of corporate culture makes it difficult to develop a universally accepted measure of corporate culture. To capture the corporate culture, some researchers have relied on tools such as surveys, questionnaires, and interviews to directly ask firm members about how they feel about the culture within their companies (Denison and Mishra, 1995; Chatman et al., 1998; Sørensen, 2002). However,

[16] The data of the score of corporate culture is available at: https://sites.google.com/view/kaili/finance-publications. Their data contains the scores of the five cultural values (innovation, integrity, quality, respect, and teamwork), as well as the number of words of conference call transcripts for each firm in a given year.

this method is subject to the significant limitation of the sample size, as well as other potential issues such as the self-selection problem. In addition, some studies develop measures of the corporate culture for a large sample of firms by using other data, such as the corporate social responsibility (CSR) data (Hoi et al., 2013) and the CEOs' past experiences (Davidson et al., 2015). But the relative accuracy and applicability of these measures is questioned. By mitigating some of the above issues, recent researchers develop new measures of corporate culture by implementing textual analysis upon different firm-related transcripts, such as the texts on the firm websites (Guiso et al., 2015b) and the textual content in the firms' 10-K reports (Audi et al., 2015; Fang et al., 2023).

In this research, I use the score of the corporate culture provided by Li et al. (2021b), who conduct the textual analysis upon the Q&A sections of public firms' conference call transcripts. As discussed in the Section 2.3.4, these new transcripts likely have several advantages over the texts that are used by the above prior studies, such as higher accuracy and spontaneity. This new culture measure concentrates on the five cultural values that are mostly discussed by S&P 500 companies: innovation, integrity, quality, respect, and teamwork. To capture the corporate culture, Li et al. (2021b) use the word embedding model on the transcripts to select words and multiword expressions that are related to the five cultural values. Such model is based on the concept that the words sharing similar neighboring words are more likely to have similar meanings. For example, if the word "collaborate" and "cooperate" tend to have common neighboring words (e.g., "share", "parties", and "partners") in the context, this suggests that these two words are more likely to have similar meanings and could be grouped together. Based on that, this method could take into account the neighboring words to obtain a numeric vector for each word. And the calculated cosine similarity between the vectors of any two words could tell whether these two words have similar meanings.

To build the culture-related word lists, Li et al. (2021b) first manually select a set of seed words for each of the five cultural values. For example, some of the seed words for the culture of quality are "customer", "dedication", "customer commitment", and "quality."

Many of these seed words for each cultural value are identical to the words that S&P 500 companies have used to describe their corresponding culture on their websites. The synonyms of these words are also added as seed words. They next use the word embedding model as the textual analysis tool to "learn" the meaning of each word through its neighboring words and represent the meaning of each word through a numeric vector. Then they select the top 500 words that tend to have similar meanings (i.e., shared neighboring words) with these seed words as the word list for each of the five cultural values. They manually check the meanings of these words in earnings calls and exclude the ones that are not suitable. Most of the removed words are named entities, industry-specific terms, or have very general meanings. Their results are robust by selecting the top 2,000 words and excluding words that appear in no more than 20 firms. In addition, if a word could be assigned to more than one cultural value, such word would be included in the word list for the cultural value with highest similarity between the word and culture seed words. In Appendix B-2, some of the culture-related words are listed as examples, together with their neighboring contents from conference call transcripts. As a result, the final word list of each cultural value contains other important words that are highly correlated with the cultural value, compared to the seed words.

After selecting the culture-related words, Li et al. (2021b) calculate the culture score as the weighted count of the number of culture-related words assigned to each cultural value scaled by total number of words in the Q&A section of each conference call transcript. The selected weighting method is the term frequency-inverse document frequency (tf-idf) weighting scheme, which calculates the product of a word's frequency within one document and the inverse document frequency of that word across documents. Specifically, the term frequency is calculated as the number of times a word appears in one document scaled by the total number of words in the document. And the inverse document frequency is calculated as the natural logarithm value of the number of all documents divided by the number of documents that include the word. Therefore, the words that are relatively more unique for the document tend to have higher values of inverse document frequency. Taken together, words that are unique to a small percentage of documents are given higher weight than words that are commonly used across all documents. For example, in a given conference call transcript, some words related to the company's innovation details tend to be more unique than other words like "company" and "financial" that might show up very frequently in this document as well as in most of the other conference call transcripts. Overall, this tf-idf weighting scheme takes account of both the importance of a word within a document and the uniqueness of such word across documents. Finally, each of the five cultural values are measured at the firm-fiscal year level by averaging the calculated culture scores from the firm conference calls within the fiscal year.

There could be some concerns about this score. One could argue that executives might conduct self-promotion during calls and this may have a significant impact on the accuracy of this method to capture whether a company has a strong culture. Thus, as a robustness test, the paragraphs with highest (top quartile) positive (negative) sentiment scores are removed from each conference call transcript. The new culture scores and the original ones are significantly correlated, ranging from 0.898 to 0.961. This indicates that executives' self-promotion does not significantly affect this measure.

Other issues could be raised about this approach. For example, it is possible that some culture-related words could be from analysts' questions during the Q&A sections, so this measure might capture something else like analysts' concerns about corporate culture. In addition, since the culture-related word lists are selected from and applied to such a great number of conference call transcripts, there is no denying that misclassification issues could occur. Moreover, the desired corporate culture can change overtime, so do the culture-related word lists. And the words and expressions might be different across firms, possibly lowering the accuracy of this measure. Thus, it is necessary to conduct validation tests to see whether the above possible problems could significantly affect the accuracy of this measure.

In response to these issues, Li et al. (2021b) conduct the validation tests and show that the score of each of the five cultural values is found to be significantly and positively correlated with the indicators of firms' corresponding practices even after controlling for firm size, operating performance, and industry and year fixed effects, which suggests good predictive validity. For example, the value of innovation culture is found to have a significantly positive relationship with all three indicators of firms' innovation activities (i.e., R&D spending, the number of patents filed and finally granted, and the dominating innovation position in the industry). Similarly, the value of respect culture is found to have a significantly positive relationship with the diversity score (the number of diversity strengths minus the number of diversity concerns) reported by KLD and the likelihood of being included in the "100 Best Companies to Work for in America" list issued by Fortune. Moreover, this measure is shown to have relatively better predictive validity than the following alternative measures, which include applying the word embedding model to the entire conference call transcripts (including the presentation and Q&A sections), using a simple count of only the seed words on Q&A sections, and implementing the word embedding model on Management's Discussion and Analysis (MD&A) section of 10-K reports.

In this study, I use their newly updated score of the corporate culture as the main dependent variable, which is provided by Professor Kai Li, one of the authors.[17] The updated score is calculated based on the earnings conference call transcripts from Thomson Reuters' StreetEvents database from 2001 to 2021 instead of the period of 2001 to 2018. Furthermore, they expand the set of culture-related word list for each cultural value from choosing the top 500 words to choosing the top 5,000 words that are most correlated with the seed words. They next use a more automated way to exclude the words that have too specific meanings in an industry or too general meanings. Specifically, they remove the words that are used by less than 20 companies from the word list. In addition, they remove the top 400 words that are most commonly used across transcripts from the word list.

To capture whether a company possess a strong corporate culture, I calculate the sum of the scores of five cultural values (i.e., innovation, integrity, quality, respect, and

[17] The data of the score of corporate culture is available at: https://sites.google.com/view/kaili/finance-publications. Their data contains the scores of the five cultural values (innovation, integrity, quality, respect, and teamwork), as well as the number of words of conference call transcripts for each firm in a given year.

teamwork). One could suggest that this total score only explains a small portion of corporate cultures, while ignoring other possible cultures (e.g., culture of safety). However, this should not be a big concern. According to the literature, these five cultural values are the most frequently discussed cultures by the S&P 500 firms (among all nine cultural values) (Guiso et al., 2015b). This indicates that these five cultural values capture a rather great part of all cultures across the S&P 500 firms which are the relatively successful companies. In addition, these five cultural values tend to have similar meanings with the six identified cultural values by Chatman et al. (2014), who group the 54 basic corporate cultural items in previous studies into six general cultural values. The following statistics results in the Section 2.5.1 suggest that the non-S&P 500 companies also think highly of these five cultural values.

2.4.2. Measuring analyst coverage and other variables

I follow the method commonly used by prior literature to measure analyst coverage (He and Tian, 2013; Guo et al., 2019; Martens and Sextroh, 2021). The relevant data of analyst earnings forecasts is collected from the I/B/E/S Summary History File. The raw measure of analyst coverage (*Coverage*) is calculated as the mean of the 12 monthly numbers of annual earnings forecasts for the firm during the one-year period before its fiscal year end. This measure is used because the majority of analysts issue at least one earnings forecast for the covering company in a given fiscal year, and most of analysts issue no more than one earnings forecast each month. Then the final measure of analyst coverage (*LnCoverage*) is calculated as the natural logarithm of one plus the raw measure of analyst coverage (*Coverage*).

An alternative method to measure analyst coverage is to count the number of analysts covering a firm in a given year. The related data is gathered from the I/B/E/S Detail History File. Following the literature (To et al., 2018; Li et al., 2019), this alternative measure is calculated as the number of unique analysts (through their unique IDs given by I/B/E/S database) that issue at least one earnings forecast for the company during the one-year period before the fiscal year end. This measure is found to be significantly correlated with the main measure (*Coverage*) in this research, and these two measures have a correlation coefficient of 0.965. Furthermore, the baseline results remain robust using this alternative measure.

As discussed above, the corporate culture could possibly be affected by firm characteristics and some of them might be correlated with the number of analysts covering the firm. Thus, it is important to control for necessary firm-related variables. However, not so much empirical research has studied and examined the determinants of corporate culture. I explore the recent literature of corporate culture (Doukas and Zhang, 2021; Hu et al., 2022; Jiang et al., 2022) to select several control variables to account for the firm characteristics that might influence a firm's culture. In particular, these control variables include: (1) firm size (*Size*), calculated as the natural logarithm value of total assets; (2) financial leverage (*Leverage*), calculated as the long-term debt plus debt in current liabilities scaled by total assets; (3) market-to-book ratio (*Market-Book ratio*); (4) sales growth (*Growth*), calculated as the percentage increase of sales from the previous year; (5) cash holdings (*Cash*), calculated as the cash and short-term investments scaled by total assets; (6) Tobin's Q (*Tobin's q*), calculated as the market value of assets divided by the book value of assets; (7) tangibility (*PPE*), calculated as the firm's property, plant and equipment scaled by total assets; and (8) capital intensity (*CAPEX*), calculated as the capital expenditure scaled by total assets. The more detailed description is shown in Table B-1. I further include the firm fixed effects and year fixed effects in the regression to control for other unobservable firm and year effects.

Furthermore, potential endogeneity issue might distort the main result. First, some factors that cannot be observed or measured might have an impact on corporate culture. If these factors are also correlated with the number of analysts covering the firms, this could distort the results in this research. In addition, the reverse causality issue may arise. One could argue that the firms with a stronger culture might attract more analysts. Thus, in later section, two identification strategies are implemented to mitigate this endogeneity concern: the two-stage least squares (2SLS) method based on an instrumental variable, and the quasi-natural experiment based on two plausibly exogenous shocks (i.e., brokerage closures and mergers).

2.5. Baseline results

2.5.1. Summary statistics

Table 2-1 lists the distribution of average scores of corporate culture and average numbers of analysts covering a firm across year (Panel A) and industry (Panel B). In Panel A, the average score of corporate culture (*TotalCulture*) ranges from 12.96 in 2009 to 17.45 in 2021. Interestingly, there is a steady increase during the period 2009 to 2021, and such trend is mostly explained by the increase in innovation (*InnovationCulture*), respect (*RespectCulture*), teamwork (*TeamworkCulture*). This could suggest that firms started to pay more attention on their culture since 2009, especially the innovation, respect, and teamwork. As for analyst coverage, the mean number of analysts (*Coverage*) covering a firm in a month range from 7.51 in 2008 to 10.24 in 2013. Over time, this number remains stable (around 8 or 9), except during the period 2007 to 2009 when firms are covered by less analysts (around 7).

Table 2-1 Sample Distribution

This table presents the mean for corporate culture-related variables and number of analysts across years (Panel A) and 12 Fama-French industries (Panel B). The 12 Fama-French industries include: (1) *NoDur* (Consumer NonDurables - Food, Tobacco, Textiles, Apparel, Leather, Toys), (2) *Durbl* (Consumer Durables - Cars, TVs, Furniture, Household Appliances), (3) *Manuf* (Manufacturing - Machinery, Trucks, Planes, Off Furn, Paper, Com Printing), (4) *Enrgy* (Energy - Oil, Gas, and Coal Extraction and Products), (5) *Chems* (Chemicals and Allied Products), (6) *BusEq* (Business Equipment - Computers, Software, and Electronic Equipment), (7) *Telcm* (Telecom - Telephone and Television Transmission), (8) *Utils* (Utilities), (9) *Shops* (Shops - Wholesale, Retail, and Some Services), (10) *Hlth* (Healthcare, Medical Equipment, and Drugs), (11) *Money* (Finance), and (12) *Other*. For definitions of these variables, please refer to Table B-1. All continuous variables are winsorized at the 1st and 99th percentiles.

Panel A: Sample distribution by year

Year	TotalCulture	InnovationCulture	IntegrityCulture	QualityCulture	RespectCulture	TeamworkCulture	Coverage
2001	14.00	4.23	2.20	2.47	3.05	2.02	8.17
2002	13.22	3.55	2.30	2.31	2.78	2.26	8.24
2003	13.74	3.74	2.35	2.43	2.88	2.31	7.89
2004	13.88	3.84	2.44	2.39	2.93	2.25	8.18
2005	14.02	3.92	2.39	2.42	2.96	2.31	8.19
2006	13.93	3.93	2.31	2.41	2.96	2.28	8.02
2007	13.72	3.91	2.27	2.39	2.88	2.23	7.88
2008	13.15	3.69	2.28	2.27	2.74	2.14	7.51
2009	12.96	3.64	2.22	2.26	2.67	2.14	7.83
2010	13.40	3.91	2.18	2.36	2.72	2.21	8.78
2011	13.78	4.15	2.17	2.43	2.76	2.22	9.22
2012	13.93	4.30	2.21	2.41	2.73	2.23	9.66
2013	14.30	4.61	2.16	2.47	2.78	2.25	10.24
2014	14.89	4.86	2.24	2.48	2.88	2.38	9.86
2015	15.08	4.95	2.25	2.39	2.99	2.44	9.67

2016	15.14	5.08	2.26	2.41	2.93	2.40	9.68
2017	15.73	5.26	2.33	2.47	3.04	2.57	9.21
2018	15.74	5.35	2.30	2.48	3.02	2.54	9.06
2019	16.20	5.61	2.31	2.53	3.14	2.56	9.10
2020	16.75	5.48	2.45	2.64	3.51	2.58	8.99
2021	17.45	6.35	2.24	3.05	3.25	2.47	9.87

Panel B: Sample distribution by industry

Industry	TotalCulture	InnovationCulture	IntegrityCulture	QualityCulture	RespectCulture	TeamworkCulture	Coverage
1: NoDur	13.41	5.43	1.98	1.81	2.40	1.75	8.42
2: Durbl	13.45	4.36	2.07	3.05	2.17	1.78	7.56
3: Manuf	12.32	3.84	1.93	2.93	1.93	1.66	7.48
4: Enrgy	10.78	3.14	2.02	2.38	1.63	1.58	13.66
5: Chems	11.54	3.70	2.10	2.24	1.85	1.64	8.90
6: BusEq	17.38	6.05	2.07	3.39	2.98	2.86	9.34
7: Telcm	16.68	5.83	2.35	2.85	2.97	2.66	9.37
8: Utils	11.33	2.60	2.76	1.91	1.88	2.15	9.29
9: Shops	14.56	5.12	2.06	2.29	3.21	1.84	9.84
10: Hlth	16.98	4.33	2.80	1.81	4.04	3.95	7.97
11: Money	12.17	3.23	2.69	1.21	3.17	1.85	8.34
12: Other	14.42	3.86	2.43	2.39	3.46	2.20	8.46

Panel B lists the distribution across the 12 Fama-French industries during the sample period. Overall, most industries think highly of innovation and have higher scores in it than other four cultural values. Intuitively, the Business Equipment (*BusEq*) industry and the Telephone and Television Transmission (*Telcm*) industry have high average scores in innovation (*InnovationCulture*). The Healthcare (*Hlth*) industry, the Utilities (*Utils*) industry, and the Finance (*Money*) industry score high in integrity (*IntegrityCulture*). The Business Equipment (*BusEq*) industry and the Consumer Durables (*Durbl*) industry score high in quality (*QualityCulture*). The Healthcare (*Hlth*) industry and the Wholesale, Retail, and Some Services (*Shops*) industry score high in respect (*RespectCulture*). The Healthcare (*Hlth*) industry, the Business Equipment (*BusEq*) industry, and the Telephone and Television Transmission (*Telcm*) industry score high in teamwork (*TeamworkCulture*). These findings based on this sample are consistent with Li et al. (2021b). Finally, the Energy industry (*Enrgy*) industry is most heavily covered by analysts.

Table 2-2 provides the summary statistics of main variables. To mitigate the impact of outliers, all continuous variables are winsorized at the 1st and 99th percentiles. Panel A reports the number of observations, mean, median, standard deviation, first quartile, and third quartile of main variables used in this empirical analysis. The results show that the number of analysts (*Coverage*) covering a firm in each month has a mean (median) value of 8.87 (6.83). Firms have an average (median) score of 14.53 (13.51) for corporate culture (*TotalCulture*). Consistent with prior results, the innovation (*InnovationCulture*) has the highest mean (median) score of 4.50 (3.88) among the five cultural values, while the integrity (*IntegrityCulture*) has the lowest mean score of 2.28 and the teamwork (*TeamworkCulture*) has the lowest median score of 1.93. Other firm characteristics have consistent patterns with prior studies. For example, the sampled firms have a mean (median) sales growth rate of 13% (7%). The tangible assets scaled by total assets have a mean (median) value of 24% (14%).

Table 2-2 Summary Statistics

This table presents the summary statistics results. Panel A reports the number of observations, mean, median, standard deviation, first quartile, and third quartile of main variables used in this empirical analysis. Panel B reports the number of observations, means (medians) for S&P 500 companies and non-S&P 500 companies. For definitions of these variables, please refer to Table B-1. All continuous variables are winsorized at the 1st and 99th percentiles. T-tests (Wilcoxon–Mann–Whitney tests) are employed to test differences between the means (medians) for S&P 500 companies and non-S&P 500 companies. *, **, and *** indicate statistical significance at the 10%, 5%, and 1% levels, respectively.

Panel A: Statistics for full sample						
Variables	N	Mean	Median	Std. dev	Q1	Q3
TotalCulture	37,263	14.53	13.51	5.50	10.46	17.56
InnovationCulture	37,263	4.50	3.88	2.59	2.65	5.71
IntegrityCulture	37,263	2.28	2.08	1.09	1.51	2.81
QualityCulture	37,263	2.43	2.09	1.46	1.37	3.11
RespectCulture	37,263	2.92	2.42	1.94	1.55	3.73
TeamworkCulture	37,263	2.34	1.93	1.53	1.29	2.93
Coverage	37,263	8.87	6.83	6.89	3.64	12.50
LnCoverage	37,263	2.05	2.06	0.71	1.53	2.60
Size	37,263	7.26	7.20	1.88	5.89	8.52
Leverage	37,263	0.22	0.20	0.19	0.04	0.35
Market-Book Ratio	37,263	3.59	2.18	4.74	1.36	3.80
Growth	37,263	0.13	0.07	0.43	-0.02	0.19
Cash	37,263	0.19	0.11	0.21	0.03	0.27
Tobins Q	37,263	2.09	1.46	4.12	1.09	2.19
PPE	37,263	0.24	0.14	0.24	0.05	0.35
CAPEX	37,263	0.04	0.03	0.05	0.01	0.05

Panel B: Statistics by groups of firms						
Variables	S&P 500			Non-S&P 500		
	N	Mean	Median	N	Mean	Median
TotalCulture	4,296	14.91	13.77	32,967	14.48***	13.48***
InnovationCulture	4,296	5.18	4.38	32,967	4.41***	3.81***
IntegrityCulture	4,296	2.29	2.11	32,967	2.28	2.07***
QualityCulture	4,296	2.35	2.02	32,967	2.44***	2.10***
RespectCulture	4,296	2.81	2.36	32,967	2.94***	2.43**
TeamworkCulture	4,296	2.25	1.96	32,967	2.36***	1.92
Coverage	4,296	15.68	15.00	32,967	7.98***	6.08***
LnCoverage	4,296	2.70	2.77	32,967	1.96***	1.96***
Size	4,296	8.70	8.67	32,967	7.08***	6.95***
Leverage	4,296	0.25	0.24	32,967	0.22***	0.19***
Market-Book Ratio	4,296	4.69	3.03	32,967	3.45***	2.10***
Growth	4,296	0.12	0.08	32,967	0.13**	0.07***
Cash	4,296	0.15	0.10	32,967	0.19***	0.11***
Tobins Q	4,296	2.16	1.65	32,967	2.08**	1.44***

PPE	4,296	0.25	0.15	32,967	0.24***	0.14***
CAPEX	4,296	0.05	0.03	32,967	0.04***	0.03***

Panel B reports the number of observations, means (medians) for S&P 500 companies and non-S&P 500 companies. The S&P 500 companies are the ones that have once been listed in the S&P 500 index during the sample period 2001 to 2021. Since the five cultural values are identified based on the culture sections from S&P 500 companies' websites, one might wonder whether these five cultural values could apply to the non-S&P 500 companies. The results show that the non-S&P 500 companies tend to be covered by less analysts, to have smaller size and lower market value. However, the non-S&P 500 companies do not seem to have a much weaker corporate culture. This finding gives confidence that the five cultural values are very likely to be emphasized by both S&P 500 and non-S&P 500 companies.

Table 2-3 reports correlations among the variables. For the dependent variables, the results show that most of the five cultural values are positively correlated with each other. Firms with higher analyst coverage tend to have a lower score of corporate culture (*TotalCulture*), which is consistent with the pressure hypothesis. Interestingly, when looking at the five cultural values separately, firms with higher analyst coverage is associated with a higher score of integrity culture, suggesting that analysts might increase the overall integrity within a firm.

Table 2-3 Correlation Matrix

This table shows correlations between variables. Variable definitions are provided in Table B-1. * indicates significance at the 1% level.

Variables	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)	(11)	(12)	(13)	(14)	(15)	(16)
(1) TotalCulture	1.000															
(2) InnovationCulture	0.787*	1.000														
(3) IntegrityCulture	0.433*	0.096*	1.000													
(4) QualityCulture	0.411*	0.277*	-0.068*	1.000												
(5) RespectCulture	0.663*	0.326*	0.331*	-0.025*	1.000											
(6) TeamworkCulture	0.665*	0.372*	0.309*	0.122*	0.321*	1.000										
(7) DocumentLength	-0.046*	0.073*	-0.021*	-0.015*	-0.087*	-0.137*	1.000									
(8) LnCoverage	-0.015*	0.141*	-0.068*	-0.036*	-0.085*	-0.093*	0.649*	1.000								
(9) Size	-0.246*	-0.076*	-0.007	-0.192*	-0.215*	-0.285*	0.481*	0.642*	1.000							
(10) Leverage	-0.146*	-0.097*	0.003	-0.072*	-0.118*	-0.141*	0.067*	0.098*	0.327*	1.000						
(11) Market-Book Ratio	0.225*	0.219*	0.034*	0.058*	0.143*	0.172*	0.075*	0.159*	-0.064*	0.152*	1.000					
(12) Growth	0.072*	0.005	0.026*	-0.007	0.071*	0.142*	-0.047*	0.003	-0.088*	-0.029*	0.104*	1.000				
(13) Cash	0.364*	0.245*	0.082*	0.122*	0.179*	0.476*	-0.091*	-0.069*	-0.438*	-0.385*	0.228*	0.160*	1.000			
(14) Tobins Q	0.152*	0.098*	0.056*	0.016*	0.089*	0.201*	-0.026*	0.017*	-0.143*	-0.084*	0.266*	0.090*	0.250*	1.000		
(15) PPE	-0.233*	-0.201*	-0.056*	0.054*	-0.225*	-0.225*	0.034*	0.101*	0.171*	0.317*	-0.078*	-0.053*	-0.356*	-0.083*	1.000	
(16) CAPEX	-0.109*	-0.076*	-0.085*	0.102*	-0.127*	-0.146*	0.055*	0.112*	-0.001	0.094*	0.006	0.050*	-0.167*	-0.019*	0.663*	1.000

One could argue that the number of analysts covering a firm could affect the length of discussion during a conference call, especially during the Q&A section, which draws attention to whether this could affect the culture measure and to what extent. First, the results from correlation matrix show a relatively large correlation coefficient (0.649) between the analyst coverage (*LnCoverage*) and the number of words in the Q&A sections of conference calls (*DocumentLength*), which supports that analyst coverage does affect the length of the Q&A section. However, the absolute values of correlation coefficients between the number of words (*DocumentLength*) and the scores of the five cultural values are quite small (from -0.137 to 0.073). This is consistent with my expectation because the measure of the five cultural values has already accounted for such effect.

Other findings show that firms with a strong corporate culture are the ones with lower size (*Size*), financial leverage (*Leverage*), tangibility (*PPE*), and capital intensity (*CAPEX*), and the ones with higher market-to-book ratio (*Market-Book ratio*), sales growth (*Growth*), cash holdings (*Cash*), and tobin's q (*Tobin's q*). In addition, firms with bigger size are covered by larger number of analysts, and these firms tend to have a longer discussion during their conference calls. Overall, these relationships could help understand the main profile of the sampled companies.

2.5.2. Impact of analyst coverage

The following model is employed to examine the impact of analyst coverage on corporate culture:

$$\begin{aligned} Culture_{j,t} = \alpha + \beta_1 LnCoverage_{j,t} + Controls + Firm\ fixed\ effects \\ + Year\ fixed\ effects + \varepsilon_{j,i,t}. \end{aligned} \tag{1}$$

The dependent variables include the corporate culture (*TotalCulture*), calculated as the sum of scores of five cultural values (i.e., innovation, integrity, quality, respect, and teamwork) for the firm in a given year, and the scores of the five constituent corporate values (*InnovationCulture*, *IntegrityCulture*, *QualityCulture*, *RespectCulture*, *TeamworkCulture*) for the firm in a given year. Since it may take time for a firm to construct its culture, I further examine whether the analyst coverage is associated with the score of corporate culture in year (t+1) to year (t+10).

The independent variable is the measure of analyst coverage (*LnCoverage*). The control variables include firm size (*Size*), financial leverage (*Leverage*), market-to-book ratio (*Market-Book ratio*), sales growth (*Growth*), cash holdings (*Cash*), Tobin's Q (*Tobin's q*), tangibility (*PPE*), capital intensity (*CAPEX*). All these measures are discussed in detail in the Section 2.4.2. The firm fixed effects and year fixed effects are added in the regression to control for other unobservable firm and year effects. The standard errors are clustered at the firm level to account for possible serial correlation of corporate culture.

Table 2-4 presents the regression results for corporate culture. The above control variables and year fixed effects are added in Column 1. The coefficient on *LnCoverage* (1.145) is significantly positive, indicating a positive relationship between analyst coverage and corporate culture. However, after adding the firm fixed effects in Column 2, the coefficient (-0.272) becomes negative and significant (at the 1% level), and the value of adjusted R-squared greatly increase (from 23.5% to 70.1%). This coefficient (-0.272) suggests that a one standard deviation increase in *LnCoverage* is associated with a 1.33% decrease (-0.272*0.71/14.53) in *TotalCulture* from its mean value. Furthermore, I find significantly negative coefficients on *LnCoverage* in Columns 3 to 9 where the dependent variables are the values of *TotalCulture* in year (t+1) to year (t+7). And these coefficients remain negative in Columns 10 to 12 where the dependent variables are the values of *TotalCulture* in year (8+1) to year (t+10).

Table 2-4 Impact of Analyst Coverage on Total Culture

This table presents the results of a model that examines the impact of analyst coverage on corporate culture. The dependent variable is *TotalCulture*, defined as the sum of a firm's five culture scores (i.e., innovation, integrity, quality, respect, and teamwork) in a given year. The independent variable is *LnCoverage*, defined as the natural logarithm of one plus the raw number of average monthly earnings forecasts. Other controls include *Size, Leverage, Market-Book Ratio, Growth, Cash, Tobins Q, PPE*, and *CAPEX*. For definitions of these variables, please refer to Table B-1. All continuous variables are winsorized at the 1st and 99th percentiles. I add firm and year fixed effects, and cluster the standard error at the firm level. T-statistics are reported in parentheses. *, **, and *** indicate statistical significance at the 10%, 5%, and 1% levels, respectively.

	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)	(11)	(12)
	TotalCulture	TotalCulture	TotalCulture (t+1)	TotalCulture (t+2)	TotalCulture (t+3)	TotalCulture (t+4)	TotalCulture (t+5)	TotalCulture (t+6)	TotalCulture (t+7)	TotalCulture (t+8)	TotalCulture (t+9)	TotalCulture (t+10)
LnCoverage	1.145***	-0.272***	-0.228**	-0.270***	-0.215**	-0.227**	-0.207**	-0.227**	-0.211*	-0.156	-0.123	-0.099
	(10.035)	(-2.848)	(-2.380)	(-2.711)	(-2.163)	(-2.279)	(-1.963)	(-2.077)	(-1.842)	(-1.299)	(-0.971)	(-0.743)
Size	-0.755***	-0.133	-0.168*	-0.119	-0.127	-0.147	-0.132	-0.069	-0.039	-0.030	-0.004	-0.014
	(-14.896)	(-1.378)	(-1.710)	(-1.196)	(-1.288)	(-1.490)	(-1.270)	(-0.613)	(-0.321)	(-0.238)	(-0.033)	(-0.100)
Leverage	-0.685**	0.507*	0.426	0.466	0.414	0.596*	0.721**	0.985***	0.949**	0.744*	0.636	0.358
	(-2.094)	(1.731)	(1.392)	(1.416)	(1.219)	(1.763)	(2.061)	(2.667)	(2.411)	(1.798)	(1.519)	(0.809)
Market-Book Ratio	0.116***	0.007	-0.002	-0.003	0.004	-0.017	-0.022*	-0.028**	-0.043***	-0.035**	-0.037***	-0.024
	(9.019)	(0.758)	(-0.157)	(-0.327)	(0.425)	(-1.464)	(-1.841)	(-2.181)	(-3.103)	(-2.442)	(-2.605)	(-1.534)
Growth	-0.011	0.072	-0.011	0.049	0.101*	0.107**	-0.003	0.093	0.072	-0.003	-0.013	0.013
	(-0.169)	(1.462)	(-0.223)	(0.951)	(1.940)	(2.050)	(-0.057)	(1.474)	(1.178)	(-0.044)	(-0.216)	(0.176)
Cash	5.179***	1.252***	0.987***	0.595*	-0.072	0.032	-0.400	-1.002***	-0.864**	-0.647	0.112	-0.088
	(14.040)	(3.564)	(2.797)	(1.671)	(-0.201)	(0.089)	(-1.077)	(-2.634)	(-2.177)	(-1.558)	(0.249)	(-0.166)
Tobins Q	0.023	0.045**	0.034	0.016	-0.050**	0.005	0.021	0.037	0.043	0.034	0.032	0.009
	(1.563)	(1.973)	(1.491)	(0.690)	(-2.376)	(0.208)	(0.836)	(1.410)	(1.397)	(0.872)	(0.871)	(0.269)
PPE	-2.722***	0.781	0.337	0.438	0.609	0.809	0.660	0.437	0.118	0.369	0.594	-0.207

	(-8.779)	(1.547)	(0.641)	(0.795)	(1.119)	(1.434)	(1.131)	(0.697)	(0.174)	(0.505)	(0.772)	(-0.257)
CAPEX	1.191	0.140	-0.505	-1.502*	-0.952	-1.200	-0.839	-0.651	-0.131	-0.528	1.330	1.955*
	(0.949)	(0.156)	(-0.546)	(-1.646)	(-1.005)	(-1.309)	(-0.900)	(-0.690)	(-0.125)	(-0.492)	(1.267)	(1.774)
Firm fixed effects	No	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Year fixed effects	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
N of observations	37263	37263	34935	32377	29973	27655	24910	22257	19873	17663	15642	13877
Adjusted R-squared	0.235	0.701	0.695	0.690	0.691	0.690	0.691	0.696	0.699	0.704	0.708	0.715

This result shows that firms covered by more analysts tend to have lower scores of the corporate culture in the current year and in the following ten years, after controlling for several important firm characteristics, which is consistent with that the analyst coverage is associated with a weaker corporate culture within the covered firms. In addition, this finding suggests that some time-invariant firm characteristics, which have noticeable effects on the corporate culture, are significantly correlated with the level of analyst coverage. This is the reason for observing a biased positive coefficient on *LnCoverage* in Column 1 and negative coefficients in Column 2 to Column 12.

As for the control variables, the findings in Columns 2 to 12 show that, after accounting for the time-invariant firm characteristics, firms with higher financial leverage (*Leverage*), higher cash holdings (*Cash*), and higher growth (*Tobin's q*) are associated with higher scores of corporate cultures, yet the coefficients on them are either not significant or significant only for a short-term period. It is also worth noting that the coefficients on *Leverage* becomes significantly positive in Columns 6 to 10 and that coefficients on *Market-Book ratio* becomes significantly negative in Columns 7 to 11, which might suggest that these two variables can have a long-term impact on the corporate culture.

As part of the robustness checks, in an unreported regression, I re-estimate the model by calculating changes in these variables. The regression results are mostly consistent with the baseline findings, supporting that the short-term pressure from analysts has a negative impact on the corporate culture. Moreover, the Difference-in-Differences model is employed in subsequent sections, which is an additional complement to this analysis.

Furthermore, in an unreported regression, the coefficients on analyst coverage remain robust after using an alternative measure of analyst coverage (i.e., the number of unique analysts following a firm in a given fiscal year). Finally, these results are consistent with the pressure hypothesis that firms with more short-term pressure from analysts are less likely to develop a set of values and norms that are widely shared and strongly held within firms.

Table 2-5 shows the regression results of the five different cultural values separately to see which cultural value is most likely to be affected by the level of analyst coverage. The dependent variables are *InnovationCulture*, *IntegrityCulture*, *QualityCulture*, *RespectCulture*, *TeamworkCulture*. Other variables are identical to those in Equation (1). For brevity, the control variables are not reported. The firm fixed effects and year fixed effects are added and the standard errors are clustered at the firm level. The results show that across all regressions, most of the coefficient estimates of *LnCoverage* are negative, consistent with the main findings in Table 2-4.

Table 2-5 Impact of Analyst Coverage on Cultural values

This table presents the results of a model that examines the impact of analyst coverage on corporate culture. The dependent variables are five cultural values (i.e., *InnovationCulture, IntegrityCulture, QualityCulture, RespectCulture, TeamworkCulture*) in a given year. The independent variable is *LnCoverage*, defined as the natural logarithm of one plus the raw number of average monthly earnings forecasts. Other controls include *Size, Leverage, Market-Book Ratio, Growth, Cash, Tobins Q, PPE*, and *CAPEX*. For definitions of these variables, please refer to Table B-1. All continuous variables are winsorized at the 1st and 99th percentiles. For brevity, the control variables are not reported. I add firm and year fixed effects, and cluster the standard error at the firm level. T-statistics are reported in parentheses. *, **, and *** indicate statistical significance at the 10%, 5%, and 1% levels, respectively.

Panel A: Innovation

	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)	(11)
	InnovationCulture	InnovationCulture (t+1)	InnovationCulture (t+2)	InnovationCulture (t+3)	InnovationCulture (t+4)	InnovationCulture (t+5)	InnovationCulture (t+6)	InnovationCulture (t+7)	InnovationCulture (t+8)	InnovationCulture (t+9)	InnovationCulture (t+10)
LnCoverage	-0.057	-0.066	-0.087*	-0.103**	-0.117**	-0.153***	-0.125**	-0.094*	-0.042	-0.023	-0.003
	(-1.248)	(-1.469)	(-1.923)	(-2.250)	(-2.523)	(-3.173)	(-2.527)	(-1.758)	(-0.746)	(-0.391)	(-0.047)
Controls	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Firm fixed effects	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Year fixed effects	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
N of observations	37263	34935	32377	29973	27655	24910	22257	19873	17663	15642	13877
Adjusted R-squared	0.718	0.717	0.713	0.716	0.716	0.719	0.724	0.726	0.728	0.729	0.736

Panel B: Integrity											
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)	(11)
	IntegrityCulture	IntegrityCulture (t+1)	IntegrityCulture (t+2)	IntegrityCulture (t+3)	IntegrityCulture (t+4)	IntegrityCulture (t+5)	IntegrityCulture (t+6)	IntegrityCulture (t+7)	IntegrityCulture (t+8)	IntegrityCulture (t+9)	IntegrityCulture (t+10)
LnCoverage	-0.060***	-0.006	-0.009	0.023	0.015	0.036	-0.001	-0.010	-0.028	-0.043	-0.044
	(-2.762)	(-0.283)	(-0.385)	(0.991)	(0.645)	(1.481)	(-0.055)	(-0.360)	(-0.966)	(-1.431)	(-1.321)
Controls	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Firm fixed effects	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Year fixed effects	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
N of observations	37263	34935	32377	29973	27655	24910	22257	19873	17663	15642	13877
Adjusted R-squared	0.488	0.483	0.486	0.488	0.487	0.492	0.502	0.508	0.516	0.521	0.520

Panel C: Quality											
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)	(11)
	QualityCulture	QualityCulture (t+1)	QualityCulture (t+2)	QualityCulture (t+3)	QualityCulture (t+4)	QualityCulture (t+5)	QualityCulture (t+6)	QualityCulture (t+7)	QualityCulture (t+8)	QualityCulture (t+9)	QualityCulture (t+10)
LnCoverage	0.019	-0.034	-0.074***	-0.078***	-0.053**	-0.033	-0.038	-0.051*	-0.028	-0.013	-0.009
	(0.810)	(-1.448)	(-3.129)	(-3.167)	(-2.133)	(-1.258)	(-1.410)	(-1.913)	(-0.978)	(-0.441)	(-0.272)
Controls	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Firm fixed effects	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Year fixed effects	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
N of observations	37263	34935	32377	29973	27655	24910	22257	19873	17663	15642	13877
Adjusted R-squared	0.696	0.695	0.692	0.687	0.687	0.689	0.692	0.697	0.702	0.710	0.714

Panel D: Respect											
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)	(11)
	RespectCulture	RespectCulture (t+1)	RespectCulture (t+2)	RespectCulture (t+3)	RespectCulture (t+4)	RespectCulture (t+5)	RespectCulture (t+6)	RespectCulture (t+7)	RespectCulture (t+8)	RespectCulture (t+9)	RespectCulture (t+10)
LnCoverage	-0.054*	-0.052*	-0.028	-0.008	-0.036	-0.010	-0.000	-0.003	-0.036	-0.017	-0.021
	(-1.775)	(-1.733)	(-0.885)	(-0.264)	(-1.135)	(-0.282)	(-0.006)	(-0.089)	(-0.946)	(-0.429)	(-0.493)
Controls	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Firm fixed effects	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Year fixed effects	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
N of observations	37263	34935	32377	29973	27655	24910	22257	19873	17663	15642	13877
Adjusted R-squared	0.689	0.684	0.684	0.684	0.686	0.686	0.690	0.691	0.696	0.704	0.710

Panel E: Teamwork											
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)	(11)
	TeamworkCulture	TeamworkCulture (t+1)	TeamworkCulture (t+2)	TeamworkCulture (t+3)	TeamworkCulture (t+4)	TeamworkCulture (t+5)	TeamworkCulture (t+6)	TeamworkCulture (t+7)	TeamworkCulture (t+8)	TeamworkCulture (t+9)	TeamworkCulture (t+10)
LnCoverage	-0.097***	-0.052**	-0.060**	-0.041	-0.024	-0.028	-0.045	-0.036	-0.002	-0.007	-0.016
	(-3.704)	(-2.042)	(-2.362)	(-1.638)	(-0.955)	(-1.013)	(-1.536)	(-1.237)	(-0.065)	(-0.216)	(-0.478)
Controls	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Firm fixed effects	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Year fixed effects	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
N of observations	37263	34935	32377	29973	27655	24910	22257	19873	17663	15642	13877
Adjusted R-squared	0.682	0.678	0.674	0.672	0.670	0.668	0.665	0.665	0.663	0.663	0.668

In particular, the coefficients on *InnovationCulture* in Panel A are negative in the relatively long-term period in Columns 3 to 8, and the absolute values of these coefficients are mostly higher than those in other Panels. This could suggest that the negative relationship between analyst coverage and the culture of innovation is stronger. I observe similar patterns of the coefficients on *QualityCulture* in Panel C, but their absolute values are relatively smaller, indicating that the negative relationships between analyst coverage and the culture of quality is not as strong as the culture of innovation. In contrast, the results in Panel B, Panel D, and Panel E show different patterns in that the negative associations between analyst coverage and culture of integrity, respect, and teamwork are negative and significant only in the relatively short-term period.

These results suggest that the short-term pressure from analysts has greater impact on certain cultural values (e.g., innovation and quality). These two cultural values tend to be long-term oriented among the five values in that these two corporate cultures represent the sets of values and norms (e.g., to encourage innovation and to promote product quality) for producing relatively long-term benefits at the cost of short-term earnings.

Overall, the above findings are consistent with the pressure hypothesis that the short-term pressure from analysts has a negative impact on the corporate culture, and the effect is more pronounced for the long-term oriented cultural values.

2.6. Identification

Although a negative relationship between analyst coverage and corporate culture is observed, the potential endogeneity issue could distort this result. For example, some omitted factors might have an impact on corporate culture and are correlated with analyst coverage. Additionally, there might be reverse causality issue, because firms with a stronger culture may attract certain analysts. To deal with endogeneity issue, two identification strategies are implemented in this section. The Section 2.6.1 introduces the two-stage least squares (2SLS) method using an instrumental variable (expected coverage). The Section 2.6.2 implements the quasi-natural experiment using two plausibly exogenous shocks (i.e., brokerage closures and mergers) on analyst coverage.

2.6.1. Instrumental variable method

To mitigate the endogeneity issue, this section uses two-stage least squares (2SLS) method based on an instrumental variable to capture the variations in analyst coverage. A perfect instrumental variable should be correlated with the level of analyst coverage but uncorrelated with the corporate culture or more specifically, the omitted variables in the residual term. Although a perfect instrumental variable has never been identified in this setting, the prior research has provided a useful instrumental variable called expected coverage (Yu, 2008). Following the literature (Yu, 2008; He and Tian, 2013; Adhikari, 2016; Kim et al., 2019; Qian et al., 2019), I calculate the change of coverage driven by the change of size of brokerage houses (i.e., number of analysts hired by the brokerage house). This instrumental variable is constructed using following equation:

$$ExpectedCoverage_{i,j,t} = (Brokersize_{j,t}/Brokersize_{j,0}) \times Coverage_{i,j,0}. \quad (2)$$

and

$$ExpectedCoverage_{j,t} = \sum_{i=1}^{n} ExpectedCoverage_{i,j,t}. \quad (3)$$

The $ExpectedCoverage_{i,j,t}$ is the expected coverage for firm j from broker i in year t. $Brokersize_{j,t}$ is the number of analysts employed by broker i in year t. $ExpectedCoverage_{j,t}$ is the sum of each broker's expected coverage of firm j in year t. The earliest year of the sample (i.e., 2001) is the benchmark year 0. The observations in the benchmark year are excluded because they are all set to one. The advantage of this instrumental variable is that the number of analysts hired by a brokerage house mostly relies on the change of brokerage revenue or earnings instead of the covered firms' corporate culture. Hence, this instrumental variable (*ExpectedCoverage*) is used to capture the part of analyst coverage (*LnCoverage*) that is almost exogenous from the corporate culture, which could show causality.

Table 2-6 shows the result. Column 1 is the first-stage regression using *LnCoverage* as dependent variable and *ExpectedCoverage* as instrumental variable. Other variables are identical to those in Equation (1). The firm fixed effects and year fixed effects are added and the standard errors are clustered at the firm level. As expected, the coefficient on *ExpectedCoverage* (0.053) is significantly positive (at 1% level) with an adjusted R-squared value of 85.8%. This indicate that the instrumental variable is significantly correlated with the independent variable (*LnCoverage*) and, thus, the coefficients on the instrumented analyst coverage in the second stage are less likely to be biased or to have a false sign.

Table 2-6 Two-Stage Least Squares Regression

This table presents the results of two-stage least squares (2SLS) regressions of corporate culture on analyst coverage. The instrumental variable, *ExpCoverage*, is defined as the sum of expected analyst coverage from all brokers, each of which is calculated as the product of the analyst coverage from a broker in benchmark year and the ratio of broker size in a given year to broker size in benchmark year. The first-stage regression predicts the value of *LnCoverage*, which is used in second-stage regression. For definitions of other variables, please refer to Table B-1. All continuous variables are winsorized at the 1st and 99th percentiles. For brevity, the control variables are not reported. I add firm and year fixed effects, and cluster the standard error at the firm level. T-statistics are reported in parentheses. *, **, and *** indicate statistical significance at the 10%, 5%, and 1% levels, respectively.

	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)	(11)	(12)
			TotalCulture	TotalCulture	TotalCulture	TotalCulture	TotalCulture	TotalCulture	TotalCulture	TotalCulture	TotalCulture	TotalCulture
	LnCoverage	TotalCulture	(t+1)	(t+2)	(t+3)	(t+4)	(t+5)	(t+6)	(t+7)	(t+8)	(t+9)	(t+10)
ExpCoverage	0.053***											
	(6.573)											
LnCoverage(predicted)		-4.445***	-5.598***	-6.069***	-6.157***	-6.097***	-6.375***	-5.642***	-4.472**	-4.197**	-3.654**	-4.082**
		(-2.677)	(-3.154)	(-3.422)	(-3.648)	(-3.603)	(-3.471)	(-3.049)	(-2.545)	(-2.445)	(-2.012)	(-2.098)
Controls	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Firm fixed effects	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Year fixed effects	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
N of observations	18745	18745	17682	16300	14989	13715	12577	11476	10389	9331	8308	7347
Adjusted R-squared	0.858	0.643	0.651	0.656	0.663	0.677	0.687	0.699	0.710	0.720	0.729	0.737

Column 2 to 12 list the results of second-stage regression. The independent variable is the predicted value of *LnCoverage* from the first-stage regression. Other settings are identical to Equation (1). Consistent with the baseline result, the coefficients on *LnCoverage(predicted)* remain significantly negative for a long-term period, with statistical significance of 1% from year 0 to year 6, and 5% from year 7 to year 10. Overall, the 2SLS results confirm the baseline results, providing evidence of causality and strengthening the pressure hypothesis that analyst coverage could have a negative impact on corporate culture.

2.6.2. Quasi-natural experiment

The second identification method uses two unexpected exogenous shocks (i.e., brokerage closures and mergers) that change the number of analysts following a firm. These events are frequently used in prior literature as quasi-natural experiments (Irani and Oesch, 2016; To et al., 2018; Li, 2020; Wang et al., 2020). Specifically, brokerage houses close their business or decide to acquire another brokerage mostly because of their own business strategy instead of the characteristics of the firms they cover. When a brokerage house is closed, its analysts would lose their jobs, decreasing the covered firms' analyst coverage. And when brokers merge, they usually fire analysts due to redundancy, especially when two analysts from these two brokers are covering the same firm (Hong and Kacperczyk, 2010). The merged brokerage house might also lose analysts because of other reasons such as status difference and culture clash among the merged entities (Wu and Zang, 2009).

These two kinds of events are first adopted by Kelly and Ljungqvist (2012) and Hong and Kacperczyk (2010), who provide the list of brokerage closures and mergers from 1984 to 2008. Fich et al. (2018) supplement this data by providing brokerage closures and mergers until 2014. In summary, the above three lists provide 77 unique exogenous events, including 26 brokerage closures and 51 brokerage mergers. Consistent with this sample period, I collect the list of these events from 2001 to 2014, which includes 59 exogenous events (23 brokerage closures and 36 brokerage mergers) fall within 2001 to 2014.

The I/B/E/S database doesn't disclose the real names of brokerage houses. It only provides the 5-digit broker codes (e.g., 01836) in I/B/E/S Detail History File and the abbreviations of brokers' real names (e.g., WITCAPTL) in I/B/E/S Recommendation File. I use the below procedure to link the broker names with the 5-digit codes in I/B/E/S Detail History File.

First, I manually compare each broker's real name in the list of brokerage closures and mergers, with the abbreviation of broker name in I/B/E/S Recommendation File to find the most possible matching (e.g., Wit Capital, Ltd. and WITCAPTL). Then I validate it by comparing the date of this broker's closure or merger with the date when I/B/E/S no longer contains any analyst recommendations from this broker. If these two dates are close to each other (e.g., 2000-05 and 2000-05-12), this pair is considered as accurate matching.

The next step is to search for the correct matching between this abbreviation of broker in I/B/E/S Recommendation File and the 5-digit broker code in I/B/E/S Detail History File. I use Python to automatically test each possible match by checking the number of mutual firms they are covering, and the date when they both stop providing forecasts or recommendations in I/B/E/S. Then the match with highest percentage of mutual firms and relatively close stop dates is considered as correct match (e.g., WITCAPTL and 01836). Following this process, the possible matches for brokerage houses (e.g., Wit Capital, Ltd., WITCAPTL, and 01836) experiencing exogenous shocks are identified. Finally, I successfully match 47 events (17 brokerage closures and 30 brokerage mergers) from 59 events.

Although the literature has identified the date information (i.e., year and month) of brokerage closures or mergers, these events might last for a relatively long period (e.g., several months). Thus, following the literature (Chen et al., 2018; Li, 2020; Bradley et al., 2022; Baruffaldi et al., 2023), the event starting (ending) date is defined as the three months before (after) the identified event date. Furthermore, the pre-event (post-event) period is defined as the one-year time before (after) the event period, which spans from 15 to 3 months before (3 to 15 months after) the specific event date. Then the analyst coverage is calculated as the natural logarithm of one plus the average number of the 12 monthly numbers of earnings forecasts for the firm in the pre- and post-event period (month -15 to -3 and 3 to 15). Other firm-year variables are calculated annually, so I use the following process to merge these variables with the event list. To avoid overlapping, I follow Chen et al. (2018) and Jing et al. (2022) to merge the firm-year observations by linking the pre-event period to the last fiscal year that ends three months before event date, and linking the post-event period to the first fiscal year that starts three months after event date. Finally, I expand the time window to include the years starting from year -3 and ending in year 12 to investigate the long-term effects[18].

Next, I follow the prior literature to identify the group of treated firms covered by the brokerage houses that experience closures or mergers during the event periods (He and Tian, 2013; Li, 2020; Baruffaldi et al., 2023). Specifically, for brokerage closures, I select the analysts who work for these brokers before the closure (month -15 to -3) and cease to cover any firm in the I/B/E/S database after the closure (month 3 to 15). The firms covered by these analysts are defined as treated firms. For brokerage mergers, I first select the firms that are covered by both target and bidding brokerage houses prior to the merger (month -15 to -3). Among these covered firms, I choose the firms that at least one of the analysts from the two brokerage houses stops covering the firm after the merger (month 3 to 15). However, I exclude the firms that are not covered by any analysts from the new brokerage house after the merger, because brokers' decision to completely stop covering a firm is likely to be endogenous. At last, these selected firms are defined as the treated firms.

Then I construct the control group based on firms that have analyst coverage in the pre-event period but are not classified as treated firms. As a robustness check, I also select the control firms that are not covered by the closing or two merging brokerage houses (Fong et al., 2022). Yet the result remains stable. Moreover, I choose the control firms that have similar firm characteristics with treated firms in the pre-event period but do not lose analyst

[18] For example, year -2 starts from month -27 to month -15, while year 2 starts from month 15 to month 27.

coverage after the event period. Following the prior literature (Derrien and KecskÉS, 2013; Chen et al., 2018; Li, 2020; Bradley et al., 2022), I require control firms to be in the same total asset quintile, Tobin's Q quintile, cash flow quintile, three-digit SIC code and year as the treated firms. I then choose the control firm that has the closest level of analyst coverage (*LnCoverage*) with the treated firm in the pre-event period. This step is to make sure the treated firms and control firms share similar characteristics which might affect corporate culture. Anyway, I control for the possible characteristics (including these matching variables) in the model discussed below.

After obtaining the matched treatment and control groups, I conduct the Difference-in-Difference (DiD) model to examine whether the exogenous shocks on analyst coverage have an impact on the corporate culture by comparing the treatment and control firms. Thus, I estimate the following model:

$$\begin{aligned} TotalCulture_{j,e,t} &= \alpha + \beta_1 Post_{e,t} + \beta_2 Treat_{j,t} + \beta_3 Post_{e,t} \times Treat_{j,t} \\ &+ Controls + Firm\ fixed\ effects + Year\ fixed\ effects \\ &+ \varepsilon_{j,e,t}.. \end{aligned} \tag{4}$$

The variable, *Post*, is a dummy variable which equals one if the firm-year observation is in the pre-event period for the event *e* and zero otherwise. *Treat* is a dummy variable which equals one if the firm is defined as a treated firm for event *e*. Other variables are identical to those in Equation (1). The firm fixed effects and year fixed effects are added. Following Chen et al. (2018) and Bradley et al. (2022), the standard errors are clustered at the event (brokerage closures or mergers) level. To mitigate the effect of extreme values, I compute the three-year average value of each of the dependent and control variables (Islam and Zein, 2020). Specifically, for each variable, I calculate the three-year average before the event period (years -1, -2, -3) and three-year average starting from year 1 (year 1, 2, 3), year 2 (year 2, 3, 4), year 3 (year 3, 4, 5) and so on until year 10 (year 10, 11, 12) after the event period. I exclude the event period (month -3 to 3) from the observations to make a distinction between the pre- and post-event period.

The results are reported in Table 2-7. From Columns 1 to 10, I examine whether the exogenous shocks on analyst coverage have an impact on corporate culture. I run this model based on different time windows including year -1 to year 1, year -1 to year 2, year -1 to year 3 and so on. Panel A reports the DiD results by using all untreated firms as control firms. Panel B reports the DiD results after selecting the control firms based on the above matched firm characteristics, which leads to a much smaller sample size. The coefficients on the interaction term between *Post* and *Treat* show the changes in treated firms' corporate culture in response to the possibly exogenous changes in analyst coverage, compared with control firms.

Table 2-7 Difference-in-differences (DiD) Results

This table presents the results of difference-in-difference regressions based on exogenous brokerage closure and merger. The dependent variable is *TotalCulture*, which is calculated separately over the pre-event period and over the post-event period. The pre- event period is defined as the three years prior to brokerage closure or merger. The post- event period is defined as the three years after brokerage closure or merger. *Treat* equals one if the firm is the treated firm for the event. *Post* equals one if the firm-year observation is in the post-event period. For definitions of other variables, please refer to Table B-1. Panel A reports the change of *TotalCulture* between treated firms and non-treated firms. Panel B reports a similar result between treated firms and matched non-treated firms. All variables are calculated as averages of annual observations over the three-year period prior to or after brokerage closure or merger. All continuous variables are winsorized at the 1st and 99th percentiles. For brevity, the control variables are not reported. I add firm and year fixed effects, and cluster the standard error at the event level. T-statistics are reported in parentheses. *, **, and *** indicate statistical significance at the 10%, 5%, and 1% levels, respectively.

Panel A: Full sample

	(1) TotalCulture (t+1)	(2) TotalCulture (t+2)	(3) TotalCulture (t+3)	(4) TotalCulture (t+4)	(5) TotalCulture (t+5)	(6) TotalCulture (t+6)	(7) TotalCulture (t+7)	(8) TotalCulture (t+8)	(9) TotalCulture (t+9)	(10) TotalCulture (t+10)
Treat*Post	0.382**	0.488**	0.588**	0.615**	0.791***	0.796***	0.839***	0.888***	0.909***	0.863***
	(2.342)	(2.069)	(2.410)	(2.393)	(3.097)	(3.060)	(3.141)	(3.222)	(3.175)	(3.063)
Post	-0.072	-0.078	-0.003	0.154	0.422*	0.738***	1.067***	1.388***	1.705***	2.038***
	(-0.857)	(-0.570)	(-0.014)	(0.679)	(1.733)	(3.053)	(4.620)	(6.660)	(8.571)	(10.747)
Treat	-0.060	-0.099	-0.145	-0.194*	-0.237**	-0.251**	-0.271**	-0.266**	-0.253**	-0.230**
	(-0.747)	(-1.003)	(-1.330)	(-1.740)	(-2.188)	(-2.390)	(-2.519)	(-2.450)	(-2.364)	(-2.111)
Controls	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Firm fixed effects	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Year fixed effects	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
N of observations	67084	64863	62900	61183	59738	58327	57060	55950	53737	52729
Adjusted R-squared	0.811	0.796	0.782	0.773	0.765	0.760	0.757	0.756	0.760	0.760

Panel B: Matched sample										
	(1) TotalCulture (t+1)	(2) TotalCulture (t+2)	(3) TotalCulture (t+3)	(4) TotalCulture (t+4)	(5) TotalCulture (t+5)	(6) TotalCulture (t+6)	(7) TotalCulture (t+7)	(8) TotalCulture (t+8)	(9) TotalCulture (t+9)	(10) TotalCulture (t+10)
Treat*Post	1.132**	1.260**	1.289**	0.849	1.056**	0.899	1.397*	1.618*	1.529*	1.504*
	(2.312)	(2.373)	(2.787)	(1.644)	(2.141)	(1.493)	(1.985)	(2.082)	(2.020)	(1.786)
Post	-0.215	0.238	0.579	1.566**	1.886***	2.195***	2.665**	2.876**	4.174***	3.979***
	(-0.347)	(0.383)	(0.880)	(2.484)	(2.934)	(2.993)	(2.370)	(2.272)	(3.330)	(3.277)
Treat	-0.349	-0.465*	-0.348	-0.739	-0.452	-0.469	-0.159	-0.509	-0.637	-0.821
	(-1.181)	(-1.924)	(-0.802)	(-1.650)	(-1.027)	(-1.213)	(-0.407)	(-1.141)	(-1.189)	(-1.462)
Controls	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Firm fixed effects	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Year fixed effects	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
N of observations	622	590	560	533	514	491	473	454	434	427
Adjusted R-squared	0.893	0.883	0.867	0.856	0.847	0.833	0.802	0.779	0.753	0.748

In Panel A, the coefficients on the interaction term (*Treat*Post*) are significantly positive (at least at 5% level) over long-term period (year 1 to year 10). Consistently, Panel B provides similar results based on the matched sample. The coefficients on the interaction term remain positive across time, though not always significant. These results show that the decrease in analyst coverage due to exogenous events (brokerage closures and mergers) could increase the score of corporate culture. This is consistent with the baseline results, supporting the pressure hypothesis.

Overall, this identification section has implemented the instrumental variable method and a quasi-natural experiment to provide further causal evidence that is consistent with the pressure hypothesis that the short-term pressure from analysts has a negative impact on the corporate culture.

2.7. Heterogeneity test

The above results are consistent with the pressure hypothesis that analysts have a negative impact on the corporate culture. In this section, I examine whether the effects of analyst coverage vary with the following analyst and firm characteristics: analyst experience, analyst forecast pressure, and corporate governance.

2.7.1. Analyst experience

Prior studies show that analysts with more experience have higher forecast accuracy (Mikhail et al., 1997; Clement, 1999; Bradley et al., 2017), provide bolder forecasts (Hong et al., 2000; Clement and Tse, 2005; Huang et al., 2017a), and incorporate more information (Mikhail et al., 2003). This might have different effects on the negative relationship between analyst coverage and corporate culture. On the one hand, experienced analysts who provide more accurate forecasts possibly impose less pressure on firm leaders. For example, if analysts make earnings forecasts that exactly equal the realized firm earnings in the near future, firm leaders are less likely to be pressured to make immediate profits to reach the target. On the other hand, with better experience, analysts could be more effective in serving their information intermediary and monitoring roles, by providing relevant firm-specific information. Therefore, analyst experience is likely to alleviate the negative impact of analyst coverage on corporate culture.

Following the literature (Clement and Tse, 2005; Yu, 2008), analyst experience is calculated based on analysts' general experience (*GenExp*) and firm-specific experience (*FirmExp*). Analyst general experience is the number of years the analyst works as an analyst, calculated as the number of years analyst provides earnings forecasts (for any firms) in the I/B/E/S Detail History File (starting from 1981). Analyst firm-specific experience is the number of years the analyst works as an analyst for the firm, calculated as the number of years analyst provides earnings forecasts for the firm in the I/B/E/S Detail History File (starting from 1981). The analyst general (firm-specific) experience is then aggregated to the firm-year level by taking the average general (firm-specific) experience of analysts following the firm in a given year.

Since the I/B/E/S database starts from 1981, the two measures of analyst experience are likely to be affected by time. For example, the measured analyst experience for firms in 1981 are automatically lower than that in 1991, because of the database start time. Thus, I scale each of the two variables annually by subtracting the minimum value of the variable in a given year from the original value, with this difference divided by the range between maximum and minimum value of the variable in that year.

The results are reported in Table 2-8. I extend Equation (1) by including the interaction term between *LnCoverage* and *High GenExp* in Columns 1 to 5, and the interaction term between *LnCoverage* and *High FirmExp* in Columns 6 to 10. *High GenExp* (*High FirmExp*) is a dummy variable, which equals one if the analyst general (firm-specific) experience is higher than the sample median. As expected, the results show that the coefficients on interaction terms are significantly positive (at 1% level) in Columns 1 to 10, indicating a mitigated negative relationship between analyst coverage and corporate culture. These results suggest that the more experienced analysts, serving better information intermediary and monitoring roles, mitigate the pressure effect of analysts on corporate culture.

Table 2-8 Effect of Analyst Experience

This table presents the results of regressions of corporate culture on analyst coverage, its interaction with measure of analyst experience, and other control variables. The dependent variable is *TotalCulture*. The independent variable is *LnCoverage*. *High GenExp* equals one if the average general experience of analysts following the firm is higher than the sample median. *High FirmExp* equals one if the average firm-specific experience of analysts following the firm is higher than the sample median. For definitions of other variables, please refer to Table B-1. All continuous variables are winsorized at the 1st and 99th percentiles. For brevity, the control variables are not reported. I add firm and year fixed effects, and cluster the standard error at the firm level. T-statistics are reported in parentheses. *, **, and *** indicate statistical significance at the 10%, 5%, and 1% levels, respectively.

	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)
	TotalCulture	TotalCulture (t+1)	TotalCulture (t+2)	TotalCulture (t+3)	TotalCulture (t+4)	TotalCulture	TotalCulture (t+1)	TotalCulture (t+2)	TotalCulture (t+3)	TotalCulture (t+4)
LnCoverage	-0.572***	-0.465***	-0.587***	-0.449***	-0.425***	-0.572***	-0.465***	-0.587***	-0.449***	-0.425***
	(-5.217)	(-4.202)	(-5.235)	(-4.041)	(-3.835)	(-5.217)	(-4.202)	(-5.235)	(-4.041)	(-3.835)
LnCoverage*High GenExp	0.500***	0.405***	0.454***	0.428***	0.450***					
	(5.310)	(4.096)	(4.558)	(4.190)	(4.215)					
High GenExp	-0.947***	-0.581***	-0.674***	-0.581***	-0.647***					
	(-4.816)	(-2.880)	(-3.312)	(-2.724)	(-2.935)					
LnCoverage*High FirmExp						0.464***	0.415***	0.516***	0.384***	0.331***
						(5.503)	(4.556)	(5.417)	(3.886)	(3.335)
High FirmExp						-0.937***	-0.678***	-0.985***	-0.825***	-0.681***
						(-5.339)	(-3.535)	(-4.959)	(-3.976)	(-3.295)
Controls	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Firm fixed effects	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Year fixed effects	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes

N of observations	36962	34621	32074	29673	27347	36962	34621	32074	29673	27347
Adjusted R-squared	0.703	0.697	0.692	0.692	0.692	0.703	0.697	0.692	0.692	0.692

2.7.2. Analyst EPS pressure

I further study the effect of analyst coverage on corporate culture in response to the short-term pressure which directly comes from analyst earnings forecast. Specifically, according to Guo et al. (2019), I calculate the difference between the firm's realized Earnings Per Share (EPS) for the fiscal year and analysts' most recent consensus forecast immediately before the firm's fiscal year end. If the firm's actual EPS is higher than the consensus forecast for EPS, firm leaders are less likely to experience short-term pressure. Therefore, the pressure effect of analyst coverage on corporate culture could be mitigated. Although this difference between the actual EPS and the consensus forecast might not possibly measure analysts' information intermediary and monitoring roles, it provides more insight into whether and how analyst coverage could impose short-term pressure.

The results are reported in Table 2-9. I extend Equation (1) by including the interaction term between *LnCoverage* and *Low Pressure*. *Low Pressure* is a dummy variable, which equals one if firms' actual EPS is higher than the most recent consensus forecasts from analysts covering the firm, and zero otherwise. The results show that the coefficients on interaction terms remain positive in Columns 1 to 5. Further, the coefficients are significant at 5% level in Columns 1, 2 and at 10% level in Column 3. This indicates that the pressure effect of analysts on corporate culture is mitigated when firm leaders are more likely to fulfil the earnings target, although such mitigation tends to be in the relatively short-term period.

Table 2-9 Effect of EPS Forecast Pressure

This table presents the results of regressions of corporate culture on analyst coverage, its interaction with measure of analysts' EPS forecast pressure, and other control variables. The dependent variable is *TotalCulture*. The independent variable is *LnCoverage*. *Low Pressure* equals one if the firm's actual EPS is lower than the most recent consensus forecast from analysts. For definitions of other variables, please refer to Table B-1. All continuous variables are winsorized at the 1st and 99th percentiles. For brevity, the control variables are not reported. I add firm and year fixed effects, and cluster the standard error at the firm level. T-statistics are reported in parentheses. *, **, and *** indicate statistical significance at the 10%, 5%, and 1% levels, respectively.

	(1)	(2)	(3)	(4)	(5)
	TotalCulture	TotalCulture (t+1)	TotalCulture (t+2)	TotalCulture (t+3)	TotalCulture (t+4)
LnCoverage	-0.354***	-0.301***	-0.327***	-0.243**	-0.260**
	(-3.452)	(-2.940)	(-3.092)	(-2.301)	(-2.453)
LnCoverage*Low Pressure	0.136**	0.123**	0.096*	0.044	0.055
	(2.425)	(2.193)	(1.650)	(0.756)	(0.883)
Low Pressure	-0.338***	-0.295**	-0.281**	-0.193	-0.146
	(-2.652)	(-2.362)	(-2.164)	(-1.507)	(-1.046)
Controls	Yes	Yes	Yes	Yes	Yes
Firm fixed effects	Yes	Yes	Yes	Yes	Yes
Year fixed effects	Yes	Yes	Yes	Yes	Yes
N of observations	37263	34935	32377	29973	27655
Adjusted R-squared	0.701	0.695	0.690	0.691	0.690

2.7.3. Corporate governance

I next investigate the impact of corporate governance on the relation between analyst coverage and corporate culture. A strong corporate governance could mitigate agency conflict, prompting firm leaders to act for the firm's interest rather than for themselves. Thus, the pressure effect of analyst coverage on corporate culture is expected to be less pronounced for firms with higher quality of corporate governance.

Following the literature (Jing et al., 2022), I use two measures of product market competition to proxy for corporate governance. The market competition is found to act as a substitute for corporate governance, which imposes discipline on executives to act for firm value (Ammann et al., 2013) and improves rational investment decisions to pursue economies of productivity and taxation (Loncan, 2022). Thus, firms in a more competitive market are expected to have better corporate governance.

To proxy for the intensity of product market competition, I employ the two text-based measures from Hoberg and Phillips (2016). They develop the Text-based Network Industry Classifications (TNIC) by calculating the textual similarity between firms' 10-k product transcripts and grouping the most similar firms with the given firm into firm-specific industries. This means that each firm has its own unique set of industry competitors. Based on this firm-specific industry classification, they calculate and provide the total product similarity between the given firm and all other firms in the same text-based industry (*TNIC Similarity*) and the Herfindahl-Hirschman Index based on the sales of the given firm and its competing firms in the text-based industry (*TNIC HHI*). A higher value of *TNIC Similarity* and a lower value of *TNIC HHI* indicate higher level of competition faced by the firm in a given year.

The results are reported in Table 2-10. I extend Equation (1) by including the interaction term between *LnCoverage* and *High TNIC Similarity* in Columns 1 to 5, and the interaction term between *LnCoverage* and *Low TNIC HHI* in Columns 6 to 10. *High TNIC Similarity* equals one if the *TNIC Similarity* is higher than the sample median, and *Low TNIC HHI* equals one if the *TNIC HHI* is lower than the sample median. The results in Columns 1 to 5 show that the coefficients on interaction terms are positive, but only the coefficients in Columns 6 and 7 are statistically significant at least at the 10% level. Though not consistently significant across time, these results suggest that the pressure effect of analyst coverage on corporate culture could be mitigated for firms with a stronger corporate governance.

Table 2-10 Effect of Corporate Governance

This table presents the results of regressions of corporate culture on analyst coverage, its interaction with measure of corporate governance, and other control variables. The dependent variable is *TotalCulture*. The independent variable is *LnCoverage*. *High TNIC Similarity* equals one if the product similarity between a given firm and all other firms in a given year is higher than the sample median. *Low TNIC HHI* equals one if the Text-based Network Industry Classifications (TNIC) Herfindahl-Hirschman Index of a given firm among the industry is lower than the sample median. For definitions of other variables, please refer to Table B-1. All continuous variables are winsorized at the 1st and 99th percentiles. For brevity, the control variables are not reported. I add firm and year fixed effects, and cluster the standard error at the firm level. T-statistics are reported in parentheses. *, **, and *** indicate statistical significance at the 10%, 5%, and 1% levels, respectively.

	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)
	TotalCulture	TotalCulture (t+1)	TotalCulture (t+2)	TotalCulture (t+3)	TotalCulture (t+4)	TotalCulture	TotalCulture (t+1)	TotalCulture (t+2)	TotalCulture (t+3)	TotalCulture (t+4)
LnCoverage	-0.329***	-0.272**	-0.283**	-0.216*	-0.244**	-0.374***	-0.395***	-0.319***	-0.253**	-0.306***
	(-2.922)	(-2.384)	(-2.337)	(-1.792)	(-2.056)	(-3.391)	(-3.544)	(-2.705)	(-2.162)	(-2.676)
LnCoverage*High TNIC Similarity	0.109	0.089	0.038	0.014	0.036					
	(0.861)	(0.703)	(0.293)	(0.104)	(0.265)					
High TNIC Similarity	-0.282	-0.114	-0.089	-0.081	-0.102					
	(-0.951)	(-0.390)	(-0.291)	(-0.265)	(-0.320)					
LnCoverage*Low TNIC HHI						0.190*	0.356***	0.124	0.086	0.174
						(1.901)	(3.517)	(1.141)	(0.796)	(1.613)
Low TNIC HHI						-0.251	-0.724***	-0.233	-0.076	-0.441*
						(-1.085)	(-3.126)	(-0.955)	(-0.318)	(-1.824)
Controls	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Firm fixed effects	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Year fixed effects	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
N of observations	35402	33271	30865	28610	26422	35347	33232	30831	28580	26398
Adjusted R-squared	0.702	0.697	0.691	0.691	0.691	0.703	0.697	0.691	0.691	0.691

2.8. Concluding remarks

In this research, I investigate the impact of analysts on the covered firms' corporate culture. The baseline results show that the firm with higher level of analyst coverage tends to have a lower score of corporate culture. In addition, the negative association is more pronounced for the constituent cultural values that tend to emphasize long-term firm value over short-term profits. These results are consistent with analysts' pressure effect that analysts may impose short-term pressure on the firm, leading to the underinvestment in corporate culture.

Furthermore, to deal with the potential endogeneity issue, I first implement the two-stage least squares (2SLS) instrumental variable method. Moreover, I conduct the quasi-natural experiment based on two events that exogenously change the analyst coverage. The results of the 2SLS regression and the Difference-in-Difference model based on the two exogenous shocks are both consistent with the observed negative association between analyst coverage and the score of corporate culture in the baseline result.

These results mostly suggest that analysts tend to have negative impact on the construction of a strong corporate culture, and this relationship is likely causal. The further results show that this negative relationship is alleviated when the firm is covered by more experienced analysts, when firm is more likely to beat analysts' earnings forecasts, when the firm has better corporate governance as captured by a more competitive market.

Taken together, this study advances the debate on the role of financial analysts in corporate governance by examining their impact on corporate culture. Although analysts are known to reduce information asymmetry and serve as external monitors, this study illustrates that the short-term pressure imposed by analysts could lead to a weaker corporate culture. This research further complements prior studies on analysts' negative impact on innovation activities by showing that analyst coverage could undermine the underlying cultural mechanisms that encourage innovation. By revealing the unintended negative consequences of analyst activities, this study emphasizes the critical role of external factors in shaping and influencing the corporate culture.

One possible limitation of this research is its use of a corporate culture measure based on textual analysis of conference call transcripts. Although this method mitigates some limitations of survey-based approaches and other textual analysis methods, its reliance on specific word lists and textual content may not fully capture the depth and complexity of corporate culture. This measure focuses on five specific cultural values (innovation, integrity, quality, respect, and teamwork) that are mostly discussed by S&P 500 companies, yet other minor cultural values are overlooked. In addition, this study is conducted in the context of U.S. markets, which might have unique attributes that could not easily generalize to other regions. For example, the cultural values and the role of financial analysts may vary significantly across different countries.

The future research could investigate the role of other external stakeholders, such as government regulators and media, in shaping corporate culture, providing a more comprehensive understanding of how external agencies affect corporate governance. It is also intriguing to extend the analysis to firms in different countries or regions to assess whether the findings are only specific to the U.S. context. Furthermore, researchers could explore alternative methods for measuring corporate culture, such as employing more advanced textual analysis methods upon firm-related materials.

2.9. Appendices

Table B-1 Variable Definitions

Variable	Description
TotalCulture	The sum of the firm's five culture scores (i.e., innovation, integrity, quality, respect, and teamwork) in a given year.
InnovationCulture	The corporate innovation culture score from Li et al. (2021b), calculated as the weighted count of the number of words related to culture of innovation (e.g., Innovation, Creativity, Excellence, Improvement, Passion) in the conference call transcripts of the firm in a given year, scaled by total number of words in the transcripts, which is aggregated to the firm-year level by taking the average.
IntegrityCulture	The corporate integrity culture score from Li et al. (2021b), calculated as the weighted count of the number of words related to culture of integrity (e.g., Integrity, Ethics, Accountability, Trust, Honesty, Responsibility) in the conference call transcripts of the firm in a given year, scaled by total number of words in the transcripts, which is aggregated to the firm-year level by taking the average.
QualityCulture	The corporate quality culture score from Li et al. (2021b), calculated as the weighted count of the number of words related to culture of quality (e.g., Quality, Customer, Meet needs, Commitment, Make a difference, Dedication) in the conference call transcripts of the firm in a given year, scaled by total number of words in the transcripts, which is aggregated to the firm-year level by taking the average.
RespectCulture	The corporate respect culture score from Li et al. (2021b), calculated as the weighted count of the number of words related to culture of respect (e.g., Respect, Diversity, Inclusion, Development, Talent, Employees, Dignity) in the conference call transcripts of the firm in a given year, scaled by total number of words in the transcripts, which is aggregated to the firm-year level by taking the average.
TeamworkCulture	The corporate teamwork culture score from Li et al. (2021b), calculated as the weighted count of the number of words related to culture of teamwork (e.g., Teamwork, Collaboration, Cooperation) in the conference call transcripts of the firm in a given year, scaled by total number of words in the transcripts, which is aggregated to the firm-year level by taking the average.
DocumentLength	The total number of words in the conference call documents for the firm in a given year, which is aggregated to the firm-year level by taking the average.
Coverage	The average number of the 12 monthly numbers of one year ahead earnings forecasts for the firm in a given year.
ExpCoverage	The sum of expected analyst coverage from all brokers covering the firm in a given year, where the expected coverage from a broker equals the product of the analyst coverage from the broker for the firm in the benchmark year and

	the ratio of broker size (i.e., the number of analysts employed by the broker) in the given year to broker size in the benchmark year.
Size	The natural logarithm value of the firm's total assets in a given year.
Leverage	The long-term debt plus the debt in current liabilities of the firm in a given year, scaled by total assets.
Market-Book Ratio	The market value of corporate equity in a given year divided by the book value of corporate equity.
Growth	The percentage increase of firm sales from the previous year.
Cash	The cash and short-term Investments of the firm in a given year, scaled by total assets.
Tobins Q	The market value of assets divided by the book value of assets, where the market value of assets equals the book value of total assets minus book value of equity plus market value of equity.
PPE	The property, plant, and equipment of the firm in a given year, scaled by total assets.
CAPEX	The capital expenditure of the firm in a given year, scaled by total assets.
GenExp	The average general experience of analysts following the firm in a given year, where the analyst general experience is calculated as the number of years an analyst has issued earnings forecasts for any firms in the I/B/E/S database.
FirmExp	The average firm experience of analysts following the firm in a given year, where the analyst firm experience is calculated as the number of years an analyst has issued earnings forecasts for the following firm.
Low Pressure	An indicator variable equal to one if firms' actual earnings is higher than the most recent consensus earnings forecasts from the analysts covering the firm, and zero otherwise.
TNIC Similarity	The measure of market competition the company faces developed by Hoberg and Phillips (2016), calculated as the sum of the cosine similarity scores between the word frequency vector of the product description in the 10-K transcript from a firm in a given year and the word frequency vectors of those from all other firms in the same year.
TNIC HHI	The measure of market power of the company developed by Hoberg and Phillips (2016), calculated as the Herfindahl-Hirschman Index based on the market shares of the given firm across all firms in the same text-based network industry, where such firm-varying industry classification includes the firm itself and other competing firms with pairwise similarities with the given firm that are above the 98th percentile in the given year..

Appendix B-2. Examples of Culture-related Words

The examples of culture-related words are listed below. The words for the cultural value of innovation includes *design*, *engineering*, *creative*, *innovate*, *environmentally friendly*, *smart*, *energy efficient*, and *exceptional*. The words for the cultural value of integrity includes *convince* and *transparency*. The words for the cultural value of quality includes *quality*, *user*, *carrier*, *ensure*, *throughput*, *upgrade*, and *customer*. The words for the cultural value of respect includes *employee*, and *office*. The words for the cultural value of teamwork includes *collaborate*, *closely*, *foster*, and *collaboration*.

OCTOBER 29, 2020 / 9:00PM, AAPL.OQ - Q4 2020 Apple Inc Earnings Call
Timothy D. Cook - Apple Inc. - CEO & Director

Yes. We're working hard to provide the best experience for our iPhone **users**. To do so, we've been **collaborating closely** with **carriers** all around the world to **ensure** iPhone has great **throughput** and coverage and battery and call **quality**. We've completed 5G testing so far on over 100 **carriers**, in over 30 regions. And so it's pretty pervasive around the world. But grantedly, it will continue to roll out in more places as carriers continue to expand their coverage, and this will happen every week. And so it's just going to get better. There are obvious places in the world where it's more ahead than in others, but we feel like we are entering at -- sort of at exactly the right time.

JANUARY 29, 2019 / 10:00PM, AAPL - Q1 2019 Apple Inc Earnings Call
Timothy D. Cook - Apple Inc. - CEO & Director

We do **design** our products to last as long as possible. Some people hold onto those for the life of the product, and some people trade them in. And then that phone is then redistributed to someone else. And so it doesn't necessarily follow that one leads to the other. The cycles -- the average cycle has extended. There's no doubt about that. We've said several times, I think, on this call and before that the **upgrades** for the quarter were less than we anticipated due to the -- all the reasons that we had mentioned. So where it goes in the future, I don't know, but I'm **convinced** that making a great product that is high **quality**, that is the best thing for the **customer** and we work for the **user**. And so that's the way that we look at it.

FEBRUARY 14, 2017 / 9:20PM, AAPL - Apple Inc at Goldman Sachs Technology and Internet Conference
Luca Maestri - Apple Inc. – CFO

The new campus, I really think, is going to be like a national landmark. I think there's going to be a lot of interest from tourists to come and visit, so we need to figure out exactly how to manage that. But this is really Steve's vision for our **employees**, particularly for the **engineering** and the **creative** teams. And his view was having an **office** space that really **fostered collaboration**. And the way the work spaces are **designed** and the layout very much encourages that type of our **collaboration**, and we think it's going to be a great place for our people to **innovate** for many, many years to come.
It's also very **environmentally friendly** -- natural ventilation, all natural light, very **smart** building controls, and very **energy efficient**, because the electricity will be generated onsite through solar and fuel cell. So I think it's going to be -- we'd be very, very happy to host you very soon.

JANUARY 31, 2017 / 10:00PM, AAPL - Q1 2017 Apple Inc Earnings Call
Tim Cook - Apple Inc. – CEO

Yes, Simona, it's Tim. We did have an **exceptional** quarter with iPhone, and that was with the back drop of not predicting the demand very well in the iPhone 7 Plus and therefore being in constraint on it through the quarter. If you look at the absolute number of upgraders, it was the highest that we've seen in any quarter. And if you look at the switcher number, it's the highest that we've seen in any quarter. If you look at the **upgrade** rate, it's similar to last year. However, I think the big asterisk, and I share all this with you for **transparency**'s sake, but I would tell you that the way we look at this is, in a quarter where you have a supply constraint, it's difficult to draw too many conclusions from it, but I wanted to share that with you anyway so you have the back drop.

3. Corporate integrity culture and analyst forecast boldness

3.1. Abstract

This study investigates whether analysts' forecasting behavior is influenced by the covered firm's information environment that is characterized by a strong integrity culture. The baseline results show that analysts tend to issue bolder earnings forecasts for the firms with higher scores of corporate integrity culture. Further results demonstrate that analysts' forecast boldness is associated with weaker market reactions for the firms with higher scores of corporate integrity culture than other firms, indicating that analysts provide relatively less informative bold opinions for firms with a stronger integrity culture. In addition, I find a negative relationship between the scores of corporate integrity culture and the number of analysts following the firm, suggesting that the analyst's service is less demanded for firms with a stronger integrity culture. Moreover, the results indicate that analysts tend to have lower forecast accuracy for firms with higher scores of corporate integrity culture. To further alleviate potential endogeneity concerns, I introduce the average score of corporate integrity culture in an industry and the CEO age of the firm as two instrumental variables to conduct the two-stage least squares (2SLS) regression. The results support the observed positive association between analyst forecast boldness and the scores of covered firms' integrity culture. Finally, I implement a robustness test by using an alternative measure of the score of corporate integrity culture based on the analyst report transcripts. The results are consistent with the above positive association, but the association is not statistically significant, probably due to the limited sample of analyst report transcripts across firms and years. Overall, the results in this research are mostly consistent with that analysts tend to have anti-herding behavior when making earnings forecasts for firms with a strong integrity culture.

3.2. Introduction

The financial analysts are regarded as one of the most important information intermediaries in the financial market, who provide analysis about the covered firms for market investors. Hence, financial analysts' behavior might be greatly affected by the corporate information environment. In this context, many prior studies have investigated the impact of corporate information environment on analysts' forecasting behavior. For example, Lehavy et al. (2011) find that the firms with more readable annual reports are associated with more accurate earnings forecasts from analysts covering these firms. Ye and Yu (2017) demonstrate that analysts tend to have less accurate forecasts for firms that restate their financial statements.

This research concentrates on whether analysts' forecasting behavior is influenced by the covered firm's information environment that is characterized by a strong integrity culture within the firm. The corporate culture indicates a set of values and norms shared within the company (O'Reilly, 1989; Van den Steen, 2010a; Fang et al., 2023). According to the prior studies (Chatman et al., 2014; Guiso et al., 2015b; Li et al., 2021b), the integrity culture that firms usually advertise tends to represent a set of values and norms that are characterized by the integrity-related attributes (e.g., integrity, trust, honesty, transparency, ethics, and accountability). Thus, the strong integrity culture suggests that the values and norms emphasizing integrity are strongly held and widely shared within the firm, which consequently guides the behavior of firm members (including firm leaders) and could lead to a corporate information environment that is characterized by the above attributes.

In this research, I investigate the impact of corporate information environment on analysts' forecasting behavior by concentrating on the forecast boldness. Other than forecast accuracy, the boldness of earnings forecasts is also found to affect the value of analyst forecasts for investors (Trueman, 1994; Gleason and Lee, 2003; Clement and Tse, 2005; Cao et al., 2023). For example, Cao et al. (2023) provide empirical findings that analysts' bolder earnings forecasts are positively associated with the short-term market-adjusted abnormal stock return. However, unlike the research into forecast accuracy, relatively fewer studies have investigated the impact of corporate information environment on analysts' forecast boldness.

According to the prior studies, the association between analysts' forecast boldness and the covered firms' integrity culture might be twofold. On the one hand, the firms with a stronger integrity culture tend to have the corporate information environment with more integrity-related attributes (e.g., higher integrity, trust, honesty, and transparency). Therefore, the better information environment of these firms is likely to reduce the information gap among analysts. Since analysts tend to have relatively more similar information set for these firms, they may issue earnings forecasts that are less deviated from other analysts. In addition, analysts might be more willing to cover these firms with better information environment. The increased potential competition among analysts might encourage them to underweight their private information and exhibit herding behavior in their forecasts. According to the above discussion, a hypothesis is reached that analysts covering firms with a stronger integrity culture tend to have less bold earnings forecasts.

On the other hand, the better corporate information environment possibly makes analysts more certain about their collected corporate information and the analysis outcome, and thus discourages analysts from taking herding behavior by mimicking others' forecasts. Moreover, the information asymmetry and the information processing are likely to be less of a problem for market investors because of this better corporate information environment, which could lead to the lower demand for analysts' service. In this case, analysts might try to provide value by collecting and analyzing new information that are not easily available for investors. In addition to analysts' efforts to collect and provide new information, analysts might take anti-herding behavior by strategically issuing forecasts that are more deviated from those of other analysts, in order to signal their ability, impress investors and generate publicity to attract clients. Taken together, this leads to an alternative hypothesis that analysts covering firms with a stronger integrity culture tend to have bolder forecasts.

To investigate the association between analysts' forecast boldness and the covered firms' integrity culture, I use the score of corporate integrity culture that is developed by Li et al. (2021b). Their method uses the word embedding model as the textual analysis tool upon the textual content in firms' conference call transcripts to identify the words that are most correlated with corporate integrity culture. Their score of the corporate integrity culture is calculated as the weighted count of the number of words that belong to their word list of the corporate integrity culture, scaled by the total number of words in the call transcript. After combining the data from different sources, the sample in this research contains the US public companies with available score of integrity culture and other data related to corporate financial information and analyst earnings forecasts, which leads to the final sample of 4,615 unique companies during the period of 2001 to 2021.

Based on this sample, I examine whether the firms' integrity culture is associated with bolder one year ahead earnings forecasts from analysts covering the firms, by regressing the measure of analysts' forecast boldness on the score of corporate integrity culture and other control variables. The model controls for a relatively large number of analyst attributes and firm characteristics that might be correlated with analysts' forecast boldness. The results of the regression model show significantly positive association between the above two variables of interest, suggesting that analysts tend to issue bolder earnings forecasts for the covered firms that are likely to have a stronger integrity culture.

Moreover, to provide more insight into the underlying explanation for such analyst forecast boldness, several further tests are conducted. First, I investigate whether the informativeness of analysts' earnings forecasts is different when analysts cover firms with a stronger integrity culture. If analysts tend issue bolder forecasts for these firms because of the discovery of new information or reduced herding behavior, these bolder forecasts should be more informative for investors. In contrast, if analysts become bolder because they tend to have anti-herding behavior by strategically differentiating themselves from others, their bold opinions are supposed to be less informative for investors than those from other analysts. The results are mostly consistent with the anti-herding behavior by showing that analysts' forecasts boldness is associated with weaker short-term investor reactions for firms with higher scores of corporate integrity culture.

Second, I investigate the association between corporate integrity culture and the number of analysts covering the firm. For one thing, analysts might be more willing to cover the firm with a stronger integrity culture due to the better corporate information environment that is characterized by more integrity-related attributes (e.g., higher integrity, trust, honesty, and transparency). On the contrary, the demand for analysts' service might be reduced since the cost of interpreting and analyzing the corporate information is likely to be lower for investors. Consistent with the latter one, the results indicate a significantly negative relationship between the score of corporate integrity culture and the number of analysts following the firm.

Third, I examine whether analysts' forecasts become less accurate when they cover the firms with a stronger integrity culture. If the information asymmetry between these firms and the analysts covering them is lower, this could lead to higher forecast accuracy for these analysts. However, if these analysts tend to issue biased forecasts due to their anti-herding behavior, they are likely to have less accurate forecasts. The regression results show that firms with a stronger integrity culture are significantly and positively associated the absolute forecast error of analysts covering the firms, consistent with the lower forecast accuracy. However, after adding the analyst-year fixed effect, this association remain positive but become insignificant, possibly because the unobserved time-varying analyst attributes could to some extent explain this relationship.

In comparison to other prior studies on analyst forecast boldness, the method in this research is less subject to potential endogeneity concerns, because the model includes a relatively larger set of control variables that capture various analyst and firm characteristics. The firm and broker fixed effects are added to control for other unobserved firm and broker attributes. The analyst-year fixed effects are also added to account for the unobserved time-variant analyst attributes. To further alleviate potential endogeneity issues, this research introduces a two-stage least squares (2SLS) instrumental variable method, using the two following instrumental variables: the average corporate integrity cultural values within an industry and the CEO age of the firm. The results from the first-stage estimation show that these two instrumental variables are significantly correlated with the score of corporate integrity culture. Using the predicted values of scores of corporate integrity culture as the new independent variable, the results from the second-stage estimation suggest that analysts are more likely to have bolder forecasts for firms with a stronger integrity culture, which is consistent with the baseline results.

As a robustness test, I introduce an alternative measure of the score of corporate integrity culture based on the analyst report transcripts. Similar to the corporate conference call transcripts, the textual content in analyst reports also contains useful information about a firm, which may provide insight into whether the firm has a strong integrity culture. Specifically, as discussed in the Section 1.4.1, the analyst report transcripts are downloaded from the Investext from the Thomson One database, which cover the S&P 500 companies from 2015 to 2020. After the processing of transcripts, I use the integrity culture-related word list from Li et al. (2021b) as the integrity culture dictionary. Similar to the previous score of corporate integrity culture, this alternative measure is calculated as the weighted count of the number of words in an analyst report transcript that appear in the integrity culture dictionary, scaled by the weighted count of all the words in the transcript, which is aggregated to the firm-year level by taking the average. Based on this alternative measure of corporate integrity culture, I find positive association between corporate integrity culture and analysts' forecast boldness, which is consistent with the baseline results. However, this association is not statistically significant, possibly due to the inadequate variation in the aggregated integrity score across firms or years.

This study makes contributions to the literature in the following ways. First, this research extends the understanding of analysts' forecasting behavior by introducing the corporate integrity culture as a fundamental and comprehensive aspect of firms' information environment. Although the prior research has examined how specific aspects of firms' information environment influence analysts' forecast boldness, these factors mostly focus on narrow aspects of firms' disclosed information. For example, the empirical findings from previous studies suggest that analysts tend to issue bolder forecasts for firms with less accessible corporate mandatory disclosures (Chang et al., 2022) and higher complexity of derivatives (Chang et al., 2016). In contrast, analysts are found to issue less bold opinions for firms with higher percentage of ownership held by the transient institutional investors as a proxy for higher corporate opaqueness (Leece and White, 2017). Moreover, analysts are found to rely more on the firms' disclosed public information and become less bold after the adoption of Regulation Fair Disclosure (Hahn and Song, 2013). In contrast, my research shifts the focus to the implication of an intrinsic and fundamental aspect of a firm's information environment, that is, the corporate culture of integrity. By showing that analysts tend to have bolder forecasts when covering firms with stronger integrity cultures, this research indicates how the firm's information environment characterized by a strong integrity culture could influence the external stakeholders' perceptions and behaviors.

Furthermore, my research provides novel insights into financial analysts' anti-herding behavior. The previous literature suggests that analysts might have the tendency to strategically differentiate themselves from the consensus, for the purpose of signaling their ability, impressing investors and generating publicity (Bernhardt et al., 2006; Ottaviani and Sorensen, 2006; Chang et al., 2016). In this context, my study finds new empirical evidence about financial analysts' anti-herding behavior when analysts are covering firms with stronger integrity cultures. Specifically, analysts' earnings forecasts for these companies are found to be relatively less informative and less accurate for investors. Consistently, other findings show that the demand for analysts' service is reduced, possibly due to the lower cost of interpreting and analyzing the corporate information for market investors, which might lead analysts to take anti-herding behavior to signal their ability and generate publicity. These findings not only deepen our understanding of the conditions under which analysts engage in anti-herding, but also highlight the unintended consequences of corporate transparency. That is, though it reduces information asymmetry for investors, it might decrease the perceived value of analysts' services, prompting them to take more aggressive forecasting strategies to maintain relevance.

Moreover, this study contributes to the literature about the impact of corporate culture. The preceding studies mostly discuss the effect of corporate culture on firm activities. For example, the empirical evidences show that a strong corporate culture is associated with faster coordination among employees and higher motivation of them (Van den Steen, 2010a), less corporate executives' short-termism (Quinn, 2018), and higher firm stability during financial crisis (Fang et al., 2023). In the context of the integrity culture, firms with a stronger integrity culture tend to have better business outcomes (i.e., productivity, profitability, industrial relations, and attractiveness to job applicants) (Guiso et al., 2015b; Graham et al., 2022a), and less accounting malfeasance (i.e., restatement) (Li et al., 2021b). Instead of the company activities, my research shifts the focus to the external effects. By examining how corporate integrity culture influences financial analysts, a key external information intermediary, this research extends the prior literature. The findings in this research suggest that corporate culture not only affects internal firm activities, but also extends its impact to the external stakeholders. Such insight emphasizes the importance of considering corporate culture as a pivotal role that influences both internal and external firm interactions.

3.3. Literature review

3.3.1. Anti-herding

Aside from analysts' herding behavior that is discussed in the Section 1.3.3, prior studies provide useful insight into analysts' anti-herding behavior that could also lead to the forecast bias. The anti-herding behavior refers to the tendency of analysts to strategically differentiate themselves from the consensus, relative to the unbiased prediction based on their own information set (Bernhardt et al., 2006; Jegadeesh and Kim, 2010; Frijns and Huynh, 2018). Based on the prior findings, analysts could have anti-herding behavior when generating their earnings forecasts (or stock recommendations) by strategically issuing forecasts (or stock recommendations) that are more deviated from those of other analysts, for the purpose of signaling their ability, impressing investors and generating publicity.

First, the prior research shows similar patterns of anti-herding behavior among other similar agents. The study on firm managers by Zwiebel (1995) finds that the firm managers with very low ability are more likely to stand out from the crowd to take more innovative actions, for the purpose of appearing more competent than other managers with average ability. In addition, Laster et al. (1999) focus on the professional forecasts issued by companies that are listed in the newsletter called *Blue Chip Economic Indicators*, where these participant firms submit their forecasts for macroeconomic variables (e.g., one year ahead real GNP/GDP growth). Consistent with the anti-herding behavior, their empirical findings suggest that the group of independent forecasting companies (e.g., consulting firms), who are more willing to seek additional clients and trade off forecast accuracy for higher publicity, tend to issue forecasts that are most deviated from the consensus. In contrast, the forecasting companies in the industrial corporation category, who need more accurate forecasts as the important input for their business operation, provide forecasts that deviate the least from the consensus. Similarly, the forecasting firms in the banks and econometric modelers groups, who need to produce accurate forecasts as an indicator to exhibit capability and credibility, also generate forecasts that are less deviated from the consensus.

Furthermore, a more recent study by Olson and Waguespack (2020) investigate another important information intermediary, the media organizations. Their study focuses on the media outlets that evaluate movies and video games on the Metacritic.com during the period of 2011 to 2014. They find that these media organizations strategically differentiate themselves from others rather than produce the objective opinions. And this strategic deviation is stronger when the critics are able to observe the opinions of others, and when they are likely to cover the same types of products. Their results are mostly consistent with that these information intermediaries tend to have anti-herding behavior in order to generate publicity among their clients.

With regard to financial analysts, Bernhardt et al. (2006) find that analysts tend to have a systematic anti-herding behavior in that analysts are more likely to issue earnings forecasts that are biased away from the consensus forecasts given the unbiased forecasts based on their own information set. In this context, Ottaviani and Sorensen (2006)'s model suggests that analysts are more encouraged to issue forecasts that are strategically deviated from the consensus forecast in a winner-take-all contest. In this competition, analysts could benefit from making unbiased forecasts conditional on the signal observed, or from strategically differentiating themselves from the consensus because the further they are from the consensus, the lower the number of analysts that would issue similar forecasts. They find that, in equilibrium, forecasters tend to move their predictions away from those from others. Clarke and Subramanian (2006) find that analysts strategically deviate their forecasts from the consensus to signal their ability, especially when their prior relative forecast performances are very poor and when they face a massive probability of losing their business.

Consistent with the above findings, Jegadeesh and Kim (2010) find that the stock market could recognize anti-herding behavior among the analysts from less famous brokerages, which is manifested by the weaker stock price reaction for analysts' recommendations that are further away from the prior consensus than those that are closer to it. That is, these analysts strategically deviate their stock recommendations from the consensus, suggesting that the analysts from less famous brokerages are likely to stand out from the crowd to attract attention. More recently, Chang et al. (2016) indicate that the complexity of corporate derivatives encourages analysts to signal their talent by relying on their expertise to make bold earnings forecasts, leading to more dispersed earnings forecasts. Moreover, Frijns and Huynh (2018) show that when analysts' covered firms receive higher attention from the media as captured by greater number of media news articles, analysts tend to generate stock recommendations that are strategically further away from the prior consensus, which is signaled by the weaker stock price reaction to the deviated recommendations by these analysts compared to those from analysts when the media coverage is lower for the covered firm. This is consistent with the anti-herding behavior among analysts that analysts are more likely to stand out from the crowd to attract attention by strategically deviating their opinions from others.

3.3.2. Corporate integrity culture

According to the above discussion in the Section 2.3.1, the prior studies propose a rather generalized definition of corporate culture that the culture of a firm represents a set of shared values and norms within the firm (O'Reilly, 1989; Kotter and Heskett, 1992; O'Reilly and Chatman, 1996). Accordingly, a strong culture represents the values and norms that are widely shared and strongly held within the organization. Such definition is also universally used by subsequent studies on corporate culture (Sørensen, 2002; Kerr and Slocum Jr, 2005; Guiso et al., 2015a; Hutton et al., 2015; Liu, 2016; Fang et al., 2023). In this context, a strong corporate integrity culture represents a set of values and norms that encourage integrity, which are widely shared and strongly held within the firm.

The prior studies have identified the culture of integrity through different methods. For example, the survey-based study by Chatman et al. (2014) have used a relatively exhaustive list of 54 different cultural values to build the profiles of the cultures of surveyed firms. After running the principal components analysis on these culture profiles, 34 of the 54 elemental cultural values are retained and grouped into six categories that represent six final cultural values (including the integrity culture). In particular, the elemental cultural values that are assigned to the integrity culture include having integrity, having high ethical standards, being honest, respecting individuals, and being fair. Similarly, Graham et al. (2022a) have implemented the survey on firm executives and ask them to describe their companies' culture. They manually label the executives' responses as reflecting firms' integrity culture if the descriptions of corporate culture in responses are associated with higher ethics, honesty, and transparency.

Apart from the survey-based method vastly used in the prior literature, other recent studies provide insight into the use of textual analysis to identify the integrity culture for a much larger sample of firms. As discussed previously, Guiso et al. (2015b) have extracted the culture-related texts from the websites of S&P 500 companies and identified nine cultural values (i.e., integrity, teamwork, innovation, respect, quality, safety, community, communication, and hard work). The integrity culture is most related to the following describing words: integrity, ethics, accountability, trust, honesty, and responsibility.

Furthermore, as discussed before, Li et al. (2021b) conduct the textual analysis model (i.e., word embedding model) on a larger sample of 9,000 public companies' conference call transcripts to select the top 5,000 words that are most associated with each of the five cultural values (innovation, integrity, quality, respect, and teamwork). This generates a much larger culture-related word list for each cultural value. In particular, the group of words that are related to the corporate integrity culture contains the words other than the six integrity-related words (i.e., integrity, ethics, accountability, trust, honesty, and responsibility) from Guiso et al. (2015b). For example, the new word list contains other important words that are highly correlated with integrity culture, such as transparency, governance, independence, objectivity, fairness, and fiduciary duty.

According to the above discussion and findings from survey-based data and firms' disclosed information, the typical integrity culture within a firm is likely associated with the following attributes, which include, but not limited to, the integrity, ethics, accountability, trust, honesty, responsibility, and transparency. Thus, it is intuitively reasonable that within the firm with a strong integrity culture, the firm members tend to hold strong shared values and norms that are correlated with the above attributes (e.g., higher integrity, trust, honesty, and transparency), which could lead to a corporate information environment that is associated with the integrity-related attributes.

Consistently, previous studies provide further empirical evidences. For example, Garrett et al. (2014) use the survey data to capture the intra-organizational trust (i.e., employees' trust in management) within firms, which could proxy for at least one of the cultural values (i.e., trust) that are associated with corporate integrity culture as discussed above. They find that such internal trust is positively associated with the financial reporting quality. Furthermore, the results from Li et al. (2017) indicate that the firms headquartered in regions of high social trust are less likely to have stock price crash risk, which is consistent with their argument that the environment with high social trust encourages firms' honest behaviors and reduces the likelihood of concealing bad news.

In addition, Jiang et al. (2017)'s results show that firms with a stronger corporate integrity culture are associated with lower investment–cash flow sensitivity. This supports their view and suggests that corporate integrity reduces the information asymmetry between stakeholders (including the credit providers) and managers, which improves companies' access to capital market and decreases firms' need to withhold cash for future investments. Moreover, Zaman (2023) has argued that a strong corporate culture, by mitigating information asymmetry, could reduce stakeholder-agency conflicts that arise due to differences in objectives. Consistently, they calculate the score of corporate culture by summing up the scores of five cultural values (i.e., innovation, integrity, quality, respect, and teamwork) and find that firms with a stronger culture are associated with less stakeholder violations. They also find similar negative association between the individual corporate integrity culture and stakeholder violations.

3.4. Hypotheses

In this research, I focus on the implications of the firm's integrity culture for one important information intermediary, the financial analysts. First, analysts might be more willing to cover firms with a strong integrity culture because these firms, as discussed above, tend to possess an information environment that is associated with the integrity-related attributes (e.g., higher integrity, trust, honesty, transparency, and accountability). The better information environment of these firms reduces the information gap among analysts and encourages them to rely more on the public information provided by these firms (e.g., during firms' earnings calls or from corporate disclosed documents). Thus, analysts are expected to make their analysis and issue their earnings forecasts based on a relatively more similar information set than those covering firms with a weak integrity culture. As suggested by prior studies, analysts tend to have similar earnings forecasts when the information is more correlated among analysts (Graham, 1999; Arya et al., 2005; Hahn and Song, 2013). This information effect suggests that analysts might appear to have herding behavior by issuing forecasts that are less deviated from the forecasts by other analysts covering the same firm, though such seemingly herding behavior does not represent analysts' biases.

In addition, since analysts might be more inclined to cover firms with a strong integrity culture, the higher number of analysts covering the same firm could intensify the competition among them, increasing their peer pressure and career concern (Wang et al., 2020). Specifically, in a more competitive environment, analysts have bigger concern of losing their business due to inaccurate analysis. According to the herding theory discussed above, analysts might choose to underweight their private information and engage in herding behavior when making earnings forecasts, by issuing forecasts that are closer to the public consensus.

In contrast to the possibly increasing herding behavior, analysts' incentives to take herding actions might be decreased for firms with a strong integrity culture. Consistent with this argument, the prior literature finds that a relatively inferior corporate information environment could lead analysts to engage in herding behavior (e.g., Frijns and Huynh (2018) and Wen and Tikoo (2020)) because these analysts are not certain about their analysis and, as mentioned above, the analysis outcome could lead to severe punishment if it deviates too much from the public consensus and turns out to be incorrect. As a result, the better information environment of firms with integrity culture makes it easier for analysts to collect and analyze the corporate information and to be more certain about their evaluation outcomes, which discourages analysts from mimicking other analysts' forecasting behavior.

Although analysts might be more willing to cover firms with a strong integrity culture, the demand for analysts' service could be decreased if analysts' information advantage over market investors is lower. The prior literature indicates that analysts could serve their information intermediary role and generate value for their clients in at least two ways: by collecting and discovering the private information that investors cannot easily access; by interpreting and analyzing the public information that cannot be easily understood or processed (Ivković and Jegadeesh, 2004; Livnat and Zhang, 2012; Huang et al., 2018). The preceding empirical evidences show that analysts' service is more required for firms with a higher information uncertainty (Lobo et al., 2012; Brown et al., 2016; Huang et al., 2018). In contrast, the firms with a stronger integrity culture tend to have better information quality (e.g., with higher integrity, trust, honesty, and transparency). This could benefit not only analysts, but also other users (e.g., institutional investors) of the corporate information by reducing the information asymmetry and information processing cost for investors.

Hence, under such circumstance, the value of information intermediaries, such as analysts, may decrease for their clients, as the cost of interpreting and analyzing the corporate information is likely to be reduced. As a result, this might encourage analysts to make more efforts to collect and analyze new information that is not easily accessible by clients, for the purpose of providing more valuable opinions for clients and attracting more investors. Thus, to the extent that analysts would engage in generating new information, analysts might issue bolder forecasts that are more deviated from the recent consensus forecast.

Furthermore, the anti-herding theory suggests that information intermediaries, such as analysts, might strategically issue forecasts that are more deviated from those of other analysts, for the purpose of signaling their ability, impressing investors and generating publicity to attract clients (Laster et al., 1999; Ottaviani and Sorensen, 2006; Chang et al., 2016; Olson and Waguespack, 2020). According to the above discussion, if analysts are more willing to cover firms with a stronger integrity culture, the higher number of analysts covering the firm would likely increase the analyst competition. Alternatively, if analysts' information advantage over market investors is lower due to the better corporate information environment, the demand for analysts' services could be reduced. In both cases, analysts might seek opportunities to act talented and generate publicity to attract more information users (e.g., institutional investors) as potential customers. Hence, they might differentiate themselves from their competitors to provide more unique opinions and produce strategically bolder forecasts.

Taken together, the above discussion suggests that firms' integrity culture could have different implications for analysts' forecasting behavior. The information effect suggests that analysts tend to possess similar information set when making forecasts for companies with a stronger integrity culture, leading analysts to have clustering forecasts. The herding theory indicates that if analysts are more inclined to cover firms with a stronger integrity culture, the possibly increased competition among analysts might encourage them to underweight their private information and exhibit herding behavior in their forecasts. Therefore, these arguments lead to the following hypothesis.

H1a: For firms with a stronger integrity culture, analysts issue less bold forecasts.

In contrast, the herding theory also suggests that the better information environment of firms with a stronger integrity culture may discourage analysts from mimicking others' forecasts. Meanwhile, the analyst effort argument demonstrates that, due to the decreasing demand for analysts' service for these firms, analysts might devote more efforts to discover and analyze new information not readily accessible to investors, focusing less on interpreting the public information that is available for most investors. Furthermore, the anti-herding theory signals that analysts covering the firms with a stronger integrity culture are likely to strategically issue forecasts that differ from those of other analysts. Thus, this discussion leads to the hypothesis below.

H1b: For firms with a stronger integrity culture, analysts issue bolder forecasts.

3.5. Sample

The following databases are used in this study. The data of sell-side analysts' forecasts is obtained from the Institutional Brokers' Estimate System (I/B/E/S) database. Specifically, the data of individual analysts' earnings forecasts is from the Detail History file, which contains analyst forecasts for annual, quarterly, semi-annual and long terms growth periods. If an analyst or broker provides more than one forecast for the firm on the same day, all forecasts are retained and provided by this database. The analysts' consensus forecast data comes from the Summary History file, which provides the summary statistical information (e.g., the mean, median, and the standard deviation) for different analysts' forecasts in a statistical period. If an analyst or broker provides more than one forecast for the firm on the same day, only the most recent forecast is used to calculate the summary statistical indicators.

According to the prior literature (Clement and Tse, 2003; Gleason and Lee, 2003; Clement and Tse, 2005; Kumar, 2010; Jiang et al., 2015; Cao et al., 2023), I collect the analysts' one year ahead forecasts for firms' annual earnings, but remove the analyst forecasts for firms that are covered by less than two unique analysts since this study compares the forecasts from different analysts. As suggested by the literature (Clement, 1999; Clement and Tse, 2003; Clement and Tse, 2005; Cao et al., 2023), I retain the analysts' annual forecasts that are issued no less than 30 days and no more than a year before the firms' fiscal year-end date, for the purpose of capturing forecasts from the active analysts.[19]

Since the measure of analysts' forecast boldness in this research depends on the calculation of the deviation of analyst forecasts from prior ones, I exclude the analyst's forecasts that have no prior forecasts by the analyst for the firm's fiscal year. Similarly, I

[19] Alternatively, I have used all forecasts without imposing this restriction. The number of observations in the baseline results increases by around 9.7%, and the results remain robust.

also remove the forecasts that have no prior year forecast data (by the analyst for the firm) for calculating the analysts' lagged forecast accuracy.

The corporate accounting data is collected from the COMPUSTAT database, while the stock market price data is obtained from the Center for Research in Security Prices (CRSP) database. The data of companies' restatement of their financial statements comes from the Audit Analytics database. The data of companies' shares held by institutional investors is collected from the Thomson Reuters' Institutional (13F) holdings database. The score of corporate integrity culture is provided by Li et al. (2021b), who construct a firm-year score for firms' integrity culture by conducting textual analysis on companies' earnings conference call transcripts from 2001 to 2021.[20] The calculation details of the variables used in this research are discussed below.

The data from the above databases are merged together. For example, the forecast data from the I/B/E/S database is merged with the corporate data from the COMPUSTAT database using the firm identifiers (i.e., the CUSIP number) and fiscal years. The observations with missing values are dropped. The final sample contains 1,001,269 individual annual earnings forecasts from 2001 to 2021.

3.5.1. Measuring forecast boldness and firm integrity culture

According to the prior studies (Gleason and Lee, 2003; Clement and Tse, 2005), an analyst's forecast is defined as a bold forecast if it is higher than both of the most recent consensus forecast and the most recent previous forecast issued by the analyst's for the firm, or else lower than both. I follow Cao et al. (2023) to construct a continuous boldness variable.[21] Specifically, the forecast boldness (*Forecast Boldness*) is calculated as the

[20] The data of the score of corporate culture is available at: https://sites.google.com/view/kaili/finance-publications. Their data contains the scores of the five cultural values: innovation, integrity, quality, respect, and teamwork. Their data also provides the number of words of conference call transcripts, which is aggregated to the firm-year level by taking average.

[21] The binary variable is more limited in cross-sectional variations.

average of the absolute deviation of an analyst's forecast from (1) the most recent consensus forecast and (2) the analyst's previous forecast, which is set to zero if the analyst's forecast is not higher or lower than both of the above two benchmarks.[22] The most recent consensus forecast is the latest previous monthly average value of one year ahead forecasts for the firm's fiscal year-end, which is available on the Summary History file of I/B/E/S database. As suggested by Clement and Tse (2005), in order to facilitate comparisons between companies, the calculated average deviation of an analyst's forecast from the consensus forecast and from the previous forecast are adjusted by the stock price two trading days before the forecast date.

With respect to the measure of corporate integrity culture, I use the score of corporate integrity culture that is introduced in detail in the Section of 2.4.1. In related to the integrity culture, Li et al. (2021b) first manually select a group of seed words for the integrity culture (e.g., integrity, accountability, ethic, honesty, transparency, and trust). Then they implement the textual analysis (i.e., the word embedding model) on the Q&A sections of conference call transcripts to study the meanings of each word through its neighboring words. Based on this method, a broader set of integrity culture-related word list is obtained by choosing the top 5,000 words that are most correlated with the above integrity seed words. Compared to the seed words, the final word list contains other important words that are highly correlated with integrity culture, such as objectivity, compliance, principled, impartial, and fiduciary duty. They next calculate the score of corporate integrity culture as the tf-idf weighted count of the number of words that are correlated with corporate integrity culture, scaled by the total number of words in the transcript. At last, the calculated scores of corporate integrity culture from call transcripts are aggregated to the firm-year level by taking the average, which is the final firm-year score of the corporate integrity culture.

In their further validation test, they examine whether the score of corporate integrity culture is negatively associated with two kinds of unethical behavior in a firm, including

[22] I do not use the peer group adjustment to adjust analyst variables to the range of zero to one, because, as suggested by Kumar (2010), such adjustment could have accounted for the systematic differences across firm-year groups and make it difficult to observe the relationship between analysts' forecast boldness and the covered firms' integrity culture.

firms' malfeasance in accounting (i.e., firms restating their financial statements) and backdating executives' option grants. Consistently, the results show that the calculated score of corporate integrity culture is found to have a significantly negative relationship with the indicators of firms' restatement of their financial statements and backdating executives' option grants. This negative relationship remains significant even after controlling for industry and year fixed effects as well as firm size and operating performance, which is consistent with good predictive validity.

3.5.2. Measuring other variables

A vast number of control variables are added in this research. I first control for several analyst characteristics that are found to be related to analysts' forecast boldness. Following the prior papers (Clement and Tse, 2005; Kumar, 2010; Jiang et al., 2015; Cleary et al., 2020; Cuculiza et al., 2020; He et al., 2020; Cao et al., 2023), I include the following control variables. Specifically, the brokerage size (*Broker Size*) is calculated as the number of analysts (who have issued earnings forecasts in the given year) from the brokerage that employs the analyst following the firm. The analyst firm-specific experience (*Firm Experience*) is defined as the number of years an analyst has issued forecasts for the firm. The analyst general experience (*General Experience*) is defined as the number of years an analyst has issued forecasts for any firms in the I/B/E/S database. The number of companies covered (*Firms Followed*) is calculated as the number of companies an analyst has issued forecasts for in a given year. The number of industries covered (*Industries Followed*) is calculated as the number of two-digit SICs an analyst has issued forecasts for in a given year. The forecast horizon (*Forecast Horizon*) is calculated as the number of days between an analyst's forecast date for the firm and the firm's fiscal year-end date. The days elapsed since last forecast (*Days Elapsed*) is calculated as the number of days between an analyst's forecast date and the date of the most recent prior forecast by any other analyst for the firm. The forecast frequency (*Forecast Frequency*) is calculated as the number of forecasts issued by an analyst for the firm in a given year. The lagged absolute forecast error (*Lag AFE*) is defined as the absolute forecast error of the last forecast issued by an analyst for the firm in the prior year, where the absolute forecast error is calculated as the absolute deviation of an analyst's forecast for the firm from the firm's actual earnings per share, scaled by the stock price two trading days before the forecast date.

Furthermore, I include the firm-level controls to account for the cross-sectional differences in firms that might be associated with analysts' forecasting behavior. Following the above literature, I first add the commonly used firm controls that are shown to be related to analysts' forecast boldness. These variables include the firm size, the market-to-book ratio, the revenue growth, the trading volume, the institutional ownership, and the stock return momentum. The firm size (*Size*) is calculated as the natural logarithm value of corporate assets. The market-to-book ratio (*Market-Book Ratio*) is calculated as the market value of corporate equity divided by the book value of corporate equity. The revenue growth (*Growth*) is calculated as the annual growth rate of firm revenue. The trading volume (*Trading Volume*) is defined as the monthly average trading volume of the firm scaled by the number of shares outstanding. The institutional ownership (*Inst Own*) is defined as the percentage of shares held by institutional investors for the firm. The stock return momentum (*ABR 12month*) is calculated as the twelve-month abnormal return prior to the forecast, where the abnormal returns is calculated as the buy-and-hold return of the stock minus the return on the value-weighted market index.

Moreover, I add other controls that are found to be associated with firms' information environment and thus affect analysts' forecasting behavior, for the purpose of alleviating the concern that the observed results might be derived from other forms of corporate information environment. Although it is not clear whether the newly added controls are related to analysts' forecast boldness based on the prior empirical evidences, they are documented by previous literature to be associated with analysts' other forecast characteristics (e.g., forecast accuracy) (Barron et al., 2008; Behn et al., 2008; Dichev and Tang, 2009; Zhang, 2010; De Franco et al., 2011; Lehavy et al., 2011; Donelson and Resutek, 2014; Ye and Yu, 2017; Yusoff et al., 2023). Following these studies, I first include several variables to proxy for different forms of firms' accounting information quality. I add the indicator of whether firms restate their financial statements (*Restatement*), which equals one if firms restate their annual or quarterly financial statements in a given year, and zero otherwise. The above findings suggest that the lower credibility of public information from restatement firms could aggravate the difficulty and complexity for analysts to make forecasts. I next add the indicator of whether firms meet or slightly beat analysts' earnings forecast (*Meet*), which equals one if a firm's actual earnings exactly meet or just beat the most preceding consensus forecast by one cent, and zero otherwise. It is found that the likelihood of firm managers meeting or slightly beating analysts' forecasts is positively related to the occurrence of myopic corporate earnings manipulation behavior. In addition, the reliability of financial information increases with audit quality. Thus, I include an indicator as the proxy for audit quality (*Auditor Quality*), which equals one if the company is audited by a Big N auditor in a given year, and zero otherwise. Moreover, I control the financial statement comparability (*Compacctind*), which is found to be related to analyst forecasting behavior in that analysts' efforts and costs are likely to be lower when analyzing the financial statements from firms with more comparable peer companies than making analysis for firms without comparable peer companies.

Furthermore, it could be more difficult for analysts to make earnings forecasts for firms that have more transitory elements in their earnings. Similar to the above research, I include the four control variables below. The earnings surprise (*Unexpected Earn*) is calculated as the absolute difference between actual corporate earnings and the earnings in last year, scaled by the stock price at the end of the last year. The negative earnings surprise (*Neg UE*) equals one if the actual corporate earnings is lower than the earnings in last year, and zero otherwise. The negative special item (*Neg SpItems*) is calculated as the absolute value of the special item scaled by the total assets if negative, and zero otherwise. The loss indicator (*Loss*) equals one if the actual corporate earnings in a given year is negative, and zero otherwise.

Next, I include two proxies to further control for the corporate information uncertainty. In particular, I follow the above literature to add the corporate earnings volatility (*Volatility*

Earn), which is calculated as the standard deviation of the firm's previous 16 quarterly earnings (scaled by total assets), and the stock return volatility (*Volatility Ret*), which is calculated as the standard deviation of the firm's previous 48 monthly stock returns. Although some of the prior control variables (e.g., *Growth*, *Restatement*, and *Unexpected Earn*) are likely to be correlated with the corporate information uncertainty, the newly added two controls could be used as more explicit proxies.

As suggested by the above research, I add the annual report readability (*Bog Index*) as a proxy for the corporate information opacity, based on the finding that the less readable corporate information in the annual reports could increase analysts' costs to interpret and process the information and thus affect analysts' forecast characteristics. The Bog Index is a reverse measure of readability that a higher value indicates less readable corporate annual report. The calculation of this index takes into account several textual attributes, including the sentence length, the English style problems, the word difficulty, and the ease of understanding the texts (Bonsall et al., 2017). This measure of readability is calculated as the natural logarithm of one plus the Bog Index of firms' 10-K reports.[23]

Moreover, I control the number of unique analysts following the firm in a year (*Analyst Cov*), calculated as the natural logarithm of one plus the number of unique analysts who have issued forecasts for the firm in a given year. According to the above studies, higher analyst coverage is associated with a better corporate information environment (e.g., better quality of corporate disclosures). The number of analysts is also correlated with the competition among analysts, which might explain analysts' forecast boldness. More importantly, the integrity culture of the firm might potentially influence the number of analysts covering the firm as discussed in the hypotheses. On the contrary, the magnitude of analyst coverage may affect the corporate culture (including the integrity culture), according to the discussion and findings in Chapter two.

In addition, I control the number of words in the Q&A section from conference call

[23] The data of the Bog Index of corporate annual reports is available at: https://kelley.iu.edu/bpm/activities/bogindex.html

transcripts (*Document Length*), calculated as the natural logarithm of one plus the average number of words in the Q&A sections of all conference call transcripts of the firm in a given year. Despite the lack of empirical findings, the word counts of Q&A sections during calls are likely associated with the number of analysts covering the firm and the magnitude of corporate disclosure, which might explain analysts' forecasting behavior. Meanwhile, the measure of corporate integrity culture is also calculated based on these particular textual materials (i.e., Q&A sections of conference call transcripts). Although the measure of corporate integrity culture is already adjusted by the number of words in the Q&A sections of conference call transcripts, I control such word counts to account for the potential correlation between it and the measure of corporate integrity culture.

Finally, I include several variables from the above literature as extra proxies for the complexity of firm business. I first control for the number of business segments of firms (*Segments*), calculated as the natural logarithm of the number of reported business segments for the firm in a given year. The literature suggests that the number of business segments is positively associated with corporate complexity and thus increases the difficulty for analysts to make earnings forecasts. Similarly, I control for the effect of corporate intangibles because the difficulty in evaluating firms' intangible assets could increase the complexity of firms' business, and thus affect analysts' forecasting behavior. Following the prior studies, I include the advertising spending (*Adv Exp*), which is defined as the advertising expense as a percentage of operating expense, and the research and development spending (*RD Exp*), which is defined as the research and development expense as a percentage of operating expense. Since many firms do not disclose these two expenses, *Adv Exp* is set to zero when the firms do not report their advertising expense. Similarly, *RD Exp* is set to zero when the firms do not report their research and development expense. And to control for the effect of such setting, I follow the above literature to add two indicator variables (*Miss_Adv Exp* and *Miss_RD Exp*), where *Miss_Adv Exp* equals one if the firm's advertising expense is missing in a given year and zero otherwise, and *Miss_RD Exp* equals one if the firm's research and development expense is missing in a given year and zero otherwise.

Moreover, to further control for other unobservable variables, several fixed effects are added. In this research, the firm and broker fixed effects are used to account for time-invariant firm and broker characteristics. In addition, the analyst fixed effect is included to control for time-invariant analyst attributes. Finally, the analyst-year fixed effect is added to control for other time-varying analyst attributes that possibly affect analysts' forecasting behavior. These specific details related to the model setting are discussed in the next section.

To alleviate the effect of outliers, all continuous variables are winsorized at the 1st and 99th percentiles. The observations with missing values for any variables are dropped. The detailed definitions of all variables are listed in Table C-1.

3.6. Baseline Results

3.6.1. Summary statistics

Table 3-1 lists the distribution of the sample by year in Panel A and by industry group in Panel B. Panel A lists the yearly distribution of the number of firms in this sample, the number of unique analysts, and the analysts' forecast revisions during the sample period. The results show that, other than the starting and ending years (2001 and 2021) in this sample, the number of analysts, their covered firms, and the number of analysts' earnings forecast revisions have been gradually increasing at a relatively steady speed. Panel B presents the distribution across the 12 Fama-French industries during the sample period.[24] The table shows that the Business Equipment (*BusEq*) industry contains the highest number of

[24] The 12 industries are respectively labelled as (1) the Consumer Nondurables (*NoDur*) industry (e.g., Food, Tobacco, Textiles, Apparel, Leather, and Toys) , (2) the Consumer Durables (*Durbl*) industry (e.g., Cars, TVs, Furniture, Household Appliances), (3) the Manufacturing (*Manuf*) industry (e.g., Machinery, Trucks, Planes, Off Furn, Paper, Com Printing), (4) the Energy (*Enrgy*) industry (e.g., Oil, Gas, and Coal Extraction and Products), (5) the Chemicals and Allied Products (*Chems*) industry, (6) the Business Equipment (*BusEq*) industry (e.g., Computers, Software, and Electronic Equipment), (7) the Telephone and Television Transmission (*Telcm*) industry, (8) the Utilities (*Utils*) industry, (9) the Wholesale, Retail, and Some Services (e.g., Laundries, Repair Shops) (*Shops*) industry, (10) the Healthcare, Medical Equipment, and Drugs (*Hlth*) industry, (11) the Finance (*Money*) industry, and (12) others (*Other*).

companies in this sample, followed by the Finance (*Money*) industry and the Healthcare, Medical Equipment, and Drugs (*Hlth*) industry. In addition, the Business Equipment (*BusEq*) industry is most heavily covered by analysts, with the highest number of unique analysts and earnings forecast revisions. Meanwhile, the Wholesale, Retail, and Some Services (*Shops*) industry, the Manufacturing (*Manuf*) industry, and the Finance (*Money*) industry are also covered by a great number of analysts, yet the number of earnings forecast revisions issued by analysts is relatively low for companies in the Manufacturing (*Manuf*) industry. In contrast, analysts are less concentrated in the Utilities (*Utils*) industry, the Telephone and Television Transmission (*Telcm*) industry, and the Chemicals and Allied Products (*Chems*) industry, with relatively lower number of unique analysts following these industries. Overall, the full sample contains 7,789 individual analysts, who follow 4,615 companies and issue 1,001,269 earnings forecast revisions during the period of 2001 to 2021.

Table 3-1 Sample Distribution

This table presents the distribution of the number of firms, unique analysts, and the analysts' forecast revisions by year (in Panel A) and by the 12 Fama-French industries (in Panel B). The 12 Fama-French industries include: (1) NoDur (Consumer NonDurables - Food, Tobacco, Textiles, Apparel, Leather, Toys), (2) Durbl (Consumer Durables - Cars, TVs, Furniture, Household Appliances), (3) Manuf (Manufacturing - Machinery, Trucks, Planes, Off Furn, Paper, Com Printing), (4) Enrgy (Energy - Oil, Gas, and Coal Extraction and Products), (5) Chems (Chemicals and Allied Products), (6) BusEq (Business Equipment - Computers, Software, and Electronic Equipment), (7) Telcm (Telecom - Telephone and Television Transmission), (8) Utils (Utilities), (9) Shops (Shops - Wholesale, Retail, and Some Services), (10) Hlth (Healthcare, Medical Equipment, and Drugs), (11) Money (Finance), and (12) Other.

Panel A: Sample distribution by year			
Year	Firms	Analysts	Forecast revisions
2001	139.00	463.00	3,691.00
2002	1,051.00	1,667.00	22,011.00
2003	1,189.00	1,768.00	26,109.00
2004	1,277.00	1,899.00	32,497.00
2005	1,480.00	2,175.00	36,887.00
2006	1,563.00	2,193.00	38,645.00
2007	1,673.00	2,236.00	41,502.00
2008	1,690.00	2,108.00	46,555.00
2009	1,653.00	2,071.00	48,440.00
2010	1,668.00	2,242.00	54,099.00
2011	1,752.00	2,390.00	57,973.00
2012	1,712.00	2,337.00	57,425.00
2013	1,737.00	2,286.00	59,914.00
2014	1,880.00	2,281.00	61,121.00
2015	1,914.00	2,264.00	65,134.00
2016	1,918.00	2,235.00	64,352.00
2017	2,029.00	2,121.00	64,878.00
2018	1,974.00	2,005.00	64,036.00
2019	1,977.00	1,992.00	63,287.00
2020	2,162.00	2,008.00	84,171.00
2021	259.00	962.00	8,542.00
Full sample	4,615.00	7,789.00	1,001,269.00
Panel B: Sample distribution by industry			
Industry	Firms	Analysts	Forecast revisions
1: NoDur	178.00	911.00	36,578.00
2: Durbl	97.00	726.00	19,464.00
3: Manuf	384.00	1,745.00	85,876.00
4: Enrgy	191.00	820.00	103,322.00
5: Chems	102.00	530.00	24,897.00

6: BusEq	1,013.00	2,964.00	198,206.00
7: Telcm	112.00	484.00	18,235.00
8: Utils	101.00	412.00	20,477.00
9: Shops	407.00	1,879.00	117,931.00
10: Hlth	683.00	1,283.00	89,877.00
11: Money	834.00	1,677.00	173,537.00
12: Other	513.00	2,545.00	112,869.00

Table 3-2 reports the summary statistics of the main variables. The statistical information (e.g., the mean, median, and standard deviation) of most variables are consistent with the prior studies. Moreover, the measure of analyst forecast boldness (*Forecast Boldness*) is multiplied by 100 to increase the readability of its statistics information. The median value of this new variable (*Forecast Boldness 100*) is higher than zero, suggesting that more than 50% of the forecast revisions are either above or below both of the most recent consensus forecasts and the previous forecasts issued by the focal analysts for the firm. In addition, on average, analysts have worked (by generating earnings forecasts) for 8.96 years as sell-side analysts in the industry and have followed each company for 5.11 years. They cover 15.23 firms from 3.39 industries in a year. They issue 6.51 one year ahead annual earnings forecasts for one firm in a year. The days between two earnings forecasts from any two analysts (*Days Elapsed*) is 6.03 days. The *Days Elapsed* has a median value of zero, indicating that more than half of the forecasts are issued on the same day right after the former forecasts. The minimum horizon of analysts' earnings forecast revisions is 30 days, consistent with the sample selection that only the forecasts issued at least 30 days before the fiscal year-end date are retained.

Table 3-2 Summary Statistics

This table presents the summary statistics results, including the number of observations, mean, median, standard deviation, first quartile, and third quartile of main variables used in this empirical analysis. For definitions of these variables, please refer to Table C-1. All continuous variables are winsorized at the 1st and 99th percentiles. *, **, and *** indicate statistical significance at the 10%, 5%, and 1% levels, respectively.

Variables	N	Mean	Median	Std. dev	Q1	Q3
Forecast Boldness 100	1,001,269	0.69	0.18	1.63	0.00	0.58
IntegrityCulture	1,001,269	2.28	2.09	1.02	1.55	2.80
Days Elapsed	1,001,269	6.03	0.00	14.43	0.00	4.00
Forecast Horizon	1,001,269	160.00	157.00	76.45	78.00	239.00
Broker Size	1,001,269	61.54	49.00	48.77	22.00	97.00
Firm Experience	1,001,269	5.11	4.00	3.33	3.00	7.00
General Experience	1,001,269	8.96	8.00	4.86	5.00	12.00
Firms Followed	1,001,269	15.23	14.00	6.78	11.00	19.00
Industries Followed	1,001,269	3.39	3.00	2.19	2.00	5.00
Forecast Frequency	1,001,269	6.51	6.00	3.05	4.00	8.00
Lag AFE	1,001,269	0.01	0.00	0.02	0.00	0.00
Size	1,001,269	8.49	8.45	1.90	7.16	9.79
Market-Book Ratio	1,001,269	3.85	2.17	5.72	1.28	3.88
Growth	1,001,269	0.10	0.07	0.27	-0.02	0.18
Trading Volume	1,001,269	0.24	0.19	0.17	0.13	0.30
Inst Own	1,001,269	0.78	0.82	0.19	0.68	0.92
ABR 12month	1,001,269	0.02	-0.02	0.38	-0.21	0.18
Restatement	1,001,269	0.10	0.00	0.30	0.00	0.00
Meet	1,001,269	0.11	0.00	0.31	0.00	0.00
Auditor Quality	1,001,269	0.93	1.00	0.26	1.00	1.00
Compacctind	1,001,269	-0.02	-0.02	0.03	-0.02	-0.01
Unexpected Earn	1,001,269	0.07	0.02	0.13	0.01	0.06
Neg UE	1,001,269	0.43	0.00	0.50	0.00	1.00
Neg SpItems	1,001,269	0.01	0.00	0.05	0.00	0.01
Loss	1,001,269	0.20	0.00	0.40	0.00	0.00
Volatility Earn	1,001,269	0.02	0.01	0.02	0.01	0.02
Volatility Ret	1,001,269	0.11	0.10	0.05	0.07	0.13
Bog Index	1,001,269	4.49	4.49	0.08	4.43	4.54
Analyst Cov	1,001,269	2.82	2.89	0.59	2.40	3.26
Document Length	1,001,269	9.30	9.37	0.61	8.92	9.79
Segments	1,001,269	0.70	0.69	0.71	0.00	1.39
Adv Exp	1,001,269	0.02	0.00	0.03	0.00	0.02
Miss_Adv Exp	1,001,269	0.55	1.00	0.50	0.00	1.00
RD Exp	1,001,269	0.07	0.00	0.13	0.00	0.07
Miss_RD Exp	1,001,269	0.45	0.00	0.50	0.00	1.00

The mean value of the score of corporate integrity culture is 2.28, which is higher than the calculated mean value (i.e., 0.584) from Li et al. (2021b). This is because my study uses the newly updated score of their corporate integrity culture, which is calculated based on the earnings conference call transcripts from 2001 to 2021. And the updated integrity culture-related word list is expanded to include the top 5,000 words (instead of 500 words) that are most correlated with the integrity seed words. After removing the unnecessary words, this word list still contains more than 2,100 words. Therefore, the newly updated score of corporate integrity culture, which is calculated as the weighted count of the number of words that are correlated with corporate integrity culture scaled by the total number of words, is likely to be higher.

3.6.2. Analyst forecast boldness

In this section, I examine whether the covered firms' integrity culture is associated with analysts' forecast boldness. The following ordinary least squares (OLS) regression model is estimated to investigate whether analysts tend to issue bolder forecasts for firms with a stronger integrity culture:

$$\begin{aligned} &Forecast\ Boldness_{i,j,t} \\ &\qquad = \alpha + \beta_1 IntegrityCulture_{j,t} + Controls + Fixed\ effects \qquad (1) \\ &\qquad + \varepsilon_{i,j,t}. \end{aligned}$$

The dependent variable, *Forecast Boldness*, captures the forecast boldness of analyst *i* covering firm *j* at time *t*. It is calculated as the average value of the absolute deviation of an analyst's forecast from (1) the most recent consensus forecast and (2) the analyst's previous forecast, which is scaled by stock price two trading days before the forecast date. The *Forecast Boldness* is set to zero if the analyst's forecast is not higher or lower than both of the above two benchmarks. If an analyst issues more than one forecast revisions on any day for the firm, then each forecast revision is treated as an observation.

The independent variable of interest, *IntegrityCulture*, is the collected score of firms' integrity culture developed by Li et al. (2021b). The *Forecast Boldness* is multiplied by 100 to improve the readability of the coefficients of variables. The T-statistics are reported in the parentheses. As discussed previously, the following control variables are added. Specifically, the analyst-level control variables include the brokerage size (*Broker Size*), the analyst firm-specific experience (*Firm Experience*), the analyst general experience (*General Experience*), the number of companies covered by analysts (*Firms Followed*), the number of industries covered by analysts (*Industries Followed*), the forecast horizon (*Forecast Horizon*), the days elapsed since last forecast (*Days Elapsed*), the forecast frequency of analysts (*Forecast Frequency*), and the lagged absolute forecast error (*Lag AFE*).

The firm-level control variables include the firm size (*Size*), the market-to-book ratio (*Market-Book Ratio*), the revenue growth (*Growth*), the trading volume (*Trading Volume*), the stock ownership held by institutional investors (*Inst Own*), the stock return momentum (*ABR 12month*), firms' restatement of their financial statements (*Restatement*), firms' meeting or slightly beating analysts' earnings forecast (*Meet*), the audit quality (*Auditor Quality*), the financial statement comparability (*Compacctind*), the earnings surprise (*Unexpected Earn*), the negative earnings surprise (*Neg UE*), the negative special item (*Neg SpItems*), firms' loss (*Loss*), the corporate earnings volatility (*Volatility Earn*), the stock return volatility (*Volatility Ret*), the annual report readability (*Bog Index*), the number of unique analysts covering the firms (*Analyst Cov*), the number of words in the Q&A sections from conference call transcripts (*Document Length*), the number of business segments of firms (*Segments*), the advertising spending (*Adv Exp* and *Miss_Adv Exp*), and the research and development spending (*RD Exp* and *Miss_RD Exp*).

I further add the firm and broker fixed effect to control for time-invariant firm and broker characteristics, as well as the year fixed effects to control for time-varying effect. The analyst fixed effect is also included to account for other time-invariant analyst characteristics (e.g., analysts' past experiences before work, gender, social convention of analysts' country of origin) that might be related to their forecasting behavior. For example, the previous studies find that analysts' forecast boldness is associated with analysts' gender (Kumar, 2010) and the individualistic culture of analysts' country of origin (Cao et al., 2023).

Moreover, to better control for other time-variant analyst attributes (e.g., analyst age, inherent ability, risk preference, income, star analyst, political preference, and overconfidence), the analyst-year fixed effect is added. For instance, the prior studies have found that analysts' forecast boldness is correlated with their individual forecast ability (Clement and Tse, 2005), political preference (Jiang et al., 2015), and risk preference (Cleary et al., 2020). Adding the analyst-year fixed effect could absorb much analyst-level heterogeneities in analysts' forecasting behavior. Finally, the standard errors are clustered at both firm and analyst level to adjust for the firm correlation and analyst correlation over time.

Table 3-3 reports the results of regressions of analysts' forecast boldness on corporate integrity culture and other control variables. The firm, broker, year fixed effects are added in Column 1, as well as the analyst fixed effect in Column 2. The analyst-year fixed effect is included in Column 3. Overall, the results (in Columns 1 to 3) show that the coefficients on the variable of interest (*IntegrityCulture*) are positive and statistically significant at the 1% level after controlling for analyst attributes and corporate characteristics. These results suggest that, for firms with a stronger integrity culture, analysts are more likely to issue bolder forecasts that are more deviated from prior consensus forecasts and their prior forecasts. The final results are mostly consistent with the hypothesis **H1b** that analysts tend to issue bolder forecasts for firms with a stronger integrity culture.[25, 26]

[25] Moreover, according to Clement and Tse (2005), I introduce an alternative measure of analyst forecast boldness, which is calculated as the absolute deviation of an analyst's forecast from the most recent consensus forecast for the same firm, scaled by stock price two trading days before the forecast date. The result is consistent with the baseline results in Table 3-3 that analysts tend to issue bolder forecasts for firms with a stronger integrity culture.

[26] As part of the robustness checks, I have re-estimated the regression model by calculating the changes in the variables across years. To do this, I retain the most recent forecast by an analyst for a firm in a given year, which reduces the sample size by nearly 83%. The regression results show that the coefficient on the change values of *IntegrityCulture* is positively significant at the 10% level, consistent with the baseline findings.

Table 3-3 Integrity Culture and Analyst Forecast Boldness

This table presents the results of regressions of analyst forecast boldness on corporate integrity culture and other control variables. The dependent variables is *Forecast Boldness*, defined as the average of the absolute difference between an analyst's current forecast and (1) the most recent consensus forecast and (2) the analyst's most recent previous forecast, scaled by two-day lagged stock price, which is set to zero if the current forecast is not higher or lower than both of the two benchmarks. *Forecast Boldness* is multiplied by 100 to improve the readability of the coefficients of variables. The independent variable is *IntegrityCulture*, defined as the score of firm's integrity culture in a given year. The analyst-level controls include *Broker Size*, *Firm Experience*, *General Experience*, *Firms Followed*, *Industries Followed*, *Forecast Horizon*, *Days Elapsed*, *Forecast Frequency*, and *Lag AFE*. The firm-level controls include *Size*, *Market-Book ratio*, *Growth, Tading Volume, Inst Own, ABR 12month, Restatement, Meet*, *Auditor Quality*, *CompAcctInd, Unexpected Earn, Neg UE, Neg SpItems*, *Loss*, *Volatility Earn*, *Volatility Ret*, *Bog Index*, *Analyst Cov, Document Length*, *Segments*, *Analyst Cov, Adv Exp, Miss_ Adv Exp, RD Exp* and *Miss_ RD Exp*. For definitions of these variables, please refer to Table C-1. All continuous variables are winsorized at the 1st and 99th percentiles. The firm, broker, year, analyst, and analyst-year fixed effects are added. The standard error is clustered at both the analyst and firm level. T-statistics are reported in parentheses. *, **, and *** indicate statistical significance at the 10%, 5%, and 1% levels, respectively.

	(1) Forecast Boldness	(2) Forecast Boldness	(3) Forecast Boldness
IntegrityCulture	0.036***	0.035***	0.026***
	(4.327)	(4.489)	(3.775)
Days Elapsed	-0.004***	-0.004***	-0.004***
	(-26.668)	(-26.772)	(-26.902)
Forecast Horizon	0.001***	0.001***	0.001***
	(14.069)	(14.232)	(14.002)
Broker Size	0.000	0.000	-0.001
	(1.126)	(0.510)	(-0.589)
Firm Experience	0.003***	0.004***	0.003***
	(3.743)	(4.068)	(3.595)
General Experience	-0.000	-0.001	
	(-0.502)	(-0.154)	
Firms Followed	0.000	0.000	
	(0.231)	(0.291)	
Industries Followed	0.002	0.008*	
	(1.138)	(1.884)	
Forecast Frequency	-0.020***	-0.016***	-0.016***
	(-13.602)	(-9.458)	(-9.408)
Lag AFE	12.661***	12.232***	12.085***
	(21.083)	(20.765)	(21.770)
Size	-0.002	-0.004	0.002

	(-0.108)	(-0.196)	(0.127)
Market-Book Ratio	-0.003*	-0.002*	-0.002*
	(-1.721)	(-1.676)	(-1.710)
Growth	-0.077**	-0.075**	-0.063**
	(-2.420)	(-2.424)	(-2.305)
Trading Volume	1.021***	1.003***	0.879***
	(8.734)	(8.972)	(9.161)
Inst Own	-0.920***	-0.931***	-0.816***
	(-8.412)	(-9.030)	(-9.493)
ABR 12month	-0.407***	-0.403***	-0.393***
	(-22.442)	(-22.602)	(-23.701)
Restatement	0.009	0.007	0.005
	(0.542)	(0.458)	(0.342)
Meet	-0.060***	-0.058***	-0.051***
	(-5.368)	(-5.341)	(-5.000)
Auditor Quality	-0.041	-0.041	-0.028
	(-0.897)	(-0.928)	(-0.679)
Compacctind	-5.283***	-5.119***	-4.577***
	(-7.053)	(-7.123)	(-6.899)
Unexpected Earn	1.012***	0.989***	0.910***
	(7.844)	(8.085)	(8.501)
Neg UE	0.047***	0.042***	0.021**
	(4.841)	(4.513)	(2.452)
Neg SpItems	0.805***	0.765***	0.770***
	(3.171)	(3.097)	(3.732)
Loss	0.375***	0.376***	0.348***
	(12.187)	(12.585)	(13.133)
Volatility Earn	-1.536***	-1.232***	-0.652
	(-3.083)	(-2.688)	(-1.554)
Volatility Ret	1.740***	2.033***	2.465***
	(5.485)	(6.588)	(8.768)
Bog Index	0.462***	0.508***	0.365***
	(2.578)	(3.145)	(2.653)
Analyst Cov	-0.334***	-0.325***	-0.292***
	(-12.084)	(-12.166)	(-11.675)
Document Length	-0.010	-0.004	-0.005
	(-0.495)	(-0.218)	(-0.315)
Segments	0.021	0.008	-0.002
	(0.787)	(0.327)	(-0.089)
Adv Exp	-0.369	-0.256	0.516
	(-0.684)	(-0.502)	(0.985)
Miss_ Adv Exp	0.046	0.047	0.064**
	(1.097)	(1.262)	(1.981)
RD Exp	-0.268	-0.216	0.022
	(-1.121)	(-0.914)	(0.103)

Miss_ RD Exp	-0.043	-0.045	-0.021
	(-1.115)	(-1.180)	(-0.587)
Firm fixed effects	Yes	Yes	Yes
Broker fixed effects	Yes	Yes	Yes
Analyst fixed effects	No	Yes	No
Year fixed effects	Yes	Yes	No
Analyst-Year fixed effects	No	No	Yes
Number of observations	1,001,269	1,001,269	1,001,269
Adjusted R-squared	0.327	0.335	0.359

The sign of coefficients on the control variables are broadly identical to the prior research. Specifically, consistent with the previous studies (Kumar, 2010; Jiang et al., 2015), the results in Table 3-3 show significantly positive coefficients on *Firm Experience*, *Forecast Horizon*, and *Lag AFE*.[27] Meanwhile, the results show significantly negative coefficients on *Days Elapsed* and *Forecast Frequency*. The coefficients on *Industries Followed* remain positive and become significant after adding the analyst fixed effect. However, the coefficients on *Broker Size, General Experience,* and *Firms Followed* are not significant. In addition, it is noted that the coefficients on *General Experience, Firms Followed,* and *Industries Followed* are automatically excluded when estimating the regression after adding the analyst-year fixed (in Column 3). This is because the analyst-year fixed effect occupies all the explanatory power of these three variables.

As for other firm-level control variables, the signs of coefficients on them are mostly consistent with the literature (Cleary et al., 2020; Cao et al., 2023). Specifically, the coefficients on *Trading Volume* are significantly positive, while the coefficients on *Market-Book ratio*, *Growth*, and *Inst Own* remain significantly negative. Additionally, the coefficients on *Size* are not significant. The only exception is the stock return momentum (*ABR 12month*), where the coefficients on *ABR 12month* are significantly negative. For the remaining controls, no prediction is made about the sign of their coefficients due to the lack of empirical evidence.

Overall, the baseline results in this section are mainly consistent with that analysts tend to issue bolder forecasts for companies with a stronger integrity culture. According to the above discussion related to hypothesis development, such boldness of analyst forecasts might reflect a lower magnitude of analyst herding behavior, analysts' better efforts to discover and analyze new information, or the increased analysts' anti-herding behavior. In the following sections, several tests are implemented to provide more insight into the source of the positive relationship between analysts' forecast boldness and the covered firms'

[27] The *Lag AFE* is a reverse measure of the analyst forecast accuracy. The significantly positive coefficient in this result is consistent with the significantly negative coefficient from the above literature.

integrity culture.

3.7. Further tests

3.7.1. Informativeness of analyst forecasts

Furthermore, I examine whether the informativeness of analysts' bold forecasts varies with the covered firms' integrity culture. If analysts issue bolder forecasts due to the discovery of new information or the decreasing herding behavior, these analysts' bold opinions should be more informative, leading to stronger market reaction to these bold forecasts. On the contrary, if analysts tend to have anti-herding behavior by strategically deviating their forecasts from others, their bold opinions should be less informative. Therefore, if analysts are more likely to have anti-herding behavior when covering firms with a stronger integrity culture, the market reaction to their bold opinions should be lower.

$$\begin{aligned} |ABR|_{j,t} = \alpha &+ \beta_1 Forecast\ Boldness_{i,j,t} + \beta_2 Strong\ IntegrityCulture_{j,t} \\ &+ \beta_3 Forecast\ Boldness_{i,j,t} * Strong\ IntegrityCulture_{j,t} \\ &+ Controls + Fixed\ effects + \varepsilon_{i,j,t}. \end{aligned} \tag{2}$$

The new dependent variable, $|ABR|$, is the absolute value of abnormal return for firm j starting from time t, where the abnormal return is calculated as the buy-and-hold return of the stock minus the return on the value-weighted market index. I calculate the absolute buy-and-hold abnormal returns based on one-, two-, and three-day time period staring from each of the forecast revision date. In particular, the $|ABR\ 0|$ is the one-day absolute value of abnormal return on day 0 (i.e., the forecast revision date). $|ABR\ 1|$ is the two-day absolute value of abnormal return from day 0 to day 1, while $|ABR\ 2|$ is the three-day absolute value of abnormal return from day 0 to day 2.

The *Forecast Boldness* is used as one of the independent variables to show whether analysts' bold forecasts are regarded as more informative. In addition, I follow Li et al. (2021b) to construct an alternative indicator of corporate integrity culture, *Strong IntegrityCulture*, which equals one if the firm's *IntegrityCulture* is among the top quartile across all firms in a year, and equals zero otherwise. Hence, the coefficient on the interaction term between *Forecast Boldness* and *Strong IntegrityCulture* could reflect whether the informativeness of analysts' bold forecasts are affected when analysts follow the firms with a stronger integrity culture.

I further add two variables (*Prior |ABR|* and *Num Forecast*) to control for other possible determinants of firms' abnormal stock return. The two control variables are correlated with the recent firm-related news. Specifically, the *Prior |ABR|* is calculated as the absolute value of the ten-day abnormal return ending the day before the forecast revision date (i.e., day -11 to day -1). The *Num Forecast* is calculated as the number of one year ahead earnings forecasts issued on the forecast revision date. Other control variables are identical to those in Eq. (1).

Table 3-4 reports the results of regressions of market reactions on analyst forecast boldness, its interaction with the indicator of corporate integrity culture (*Strong IntegrityCulture*), and other control variables. The dependent variable, *|ABR|*, is measured respectively over the one-, two-, three-day time window starting from the analyst forecast date.

Table 3-4 Investor Reactions to Analyst Forecast

This table presents the results of regressions of investor reactions on analyst forecast boldness, its interaction with integrity culture, and other control variables. The dependent variables is |*ABR*|, defined as the absolute value of abnormal return staring from the forecast revision date, where the abnormal return is calculated as the buy-and-hold return of the stock minus the buy-and-hold return on the value-weighted market index. The |*ABR*| is calculated for one-, two-, and three-day period. *Forecast Boldness* is defined as the average of the absolute difference between an analyst's current forecast and (1) the most recent consensus forecast and (2) the analyst's most recent previous forecast, which is set to zero if the current forecast is not higher or lower than both of the two benchmarks. *Strong IntegrityCulture* equals one if the firm's *IntegrityCulture* is among the top quartile across all firms in a given year, and zero otherwise. The analyst-level controls include *Broker Size*, *Firm Experience*, *General Experience*, *Firms Followed*, *Industries Followed*, *Forecast Horizon*, *Days Elapsed*, *Forecast Frequency*, and *Lag AFE*. The firm-level controls include *Size*, *Market-Book ratio*, *Growth, Tading Volume, Inst Own, ABR 12month, Restatement, Meet*, *Auditor Quality*, *CompAcctInd, Unexpected Earn, Neg UE, Neg SpItems*, *Loss*, *Volatility Earn*, *Volatility Ret*, *Bog Index*, *Analyst Cov, Document Length*, *Segments*, *Analyst Cov, Adv Exp, Miss_ Adv Exp, RD Exp* and *Miss_ RD Exp*, *Prior* |*ABR*|, and *Num Forecast*. For definitions of these variables, please refer to Table C-1. All continuous variables are winsorized at the 1st and 99th percentiles. The firm, broker, and analyst-year fixed effects are added. For brevity, analyst-level and firm-level controls are not reported. The standard error is clustered at both the analyst and firm level. T-statistics are reported in parentheses. *, **, and *** indicate statistical significance at the 10%, 5%, and 1% levels, respectively.

	(1)	(2)	(3)
	|ABR 0|	|ABR 1|	|ABR 2|
Forecast Boldness	0.237***	0.314***	0.338***
	(20.156)	(21.105)	(20.435)
Strong IntegrityCulture	0.000	0.000	0.001
	(0.589)	(0.694)	(1.421)
Forecast Boldness*Strong IntegrityCulture	-0.063***	-0.076***	-0.075***
	(-4.293)	(-3.892)	(-3.346)
Analyst Controls	Yes	Yes	Yes
Firm Controls	Yes	Yes	Yes
Firm fixed effects	Yes	Yes	Yes
Broker fixed effects	Yes	Yes	Yes
Analyst-Year fixed effects	Yes	Yes	Yes
Number of observations	1,001,269	1,001,269	1,001,269
Adjusted R-squared	0.302	0.295	0.285

The coefficients on *Forecast Boldness* are significant and positive at the 1% level (in Columns 1 through 3), indicating that analysts' bold forecast revisions are more informative to the market investors. Nevertheless, the coefficients on the interaction term between *Forecast Boldness* and *Strong IntegrityCulture* are negative and statistically significant at the 1% level, suggesting that market discounts the bold forecasts from analysts who cover firms with a stronger integrity culture. This is consistent with analysts' anti-herding behavior that analysts strategically deviate their forecasts from others, which reduces the informativeness of their forecasts.

Overall, the results in Table 3-4 are more consistent with analysts' anti-herding behavior that the bold forecasts from analysts who cover firms with a stronger integrity culture are less informative, possibly because these analysts strategically deviate their forecasts from others regardless of providing new information to the market.

3.7.2. Analyst coverage of firms

In this section, I examine the association between corporate integrity culture and the number of analysts covering the firm. As discussed in the hypothesis part, analysts might have more incentives to cover firms with a stronger integrity culture because of a better corporate information environment (e.g., with higher integrity, trust, honesty, and transparency). On the contrary, the demand for analysts' service might be decreased due to the reduced information asymmetry and information processing cost for investors. To examine whether and how the covered firms' integrity culture is associated with number of analysts covering the firms, the following OLS regression model is implemented:

$$Analyst\ Cov_{j,t} = \alpha + \beta_1 IntegrityCulture_{j,t} + Controls + Fixed\ effects + \varepsilon_{j,t}. \quad (3)$$

The *Analyst Cov* from Eq. (1) is used as the new dependent variable, which is calculated as the natural logarithm of one plus the number of unique analysts who have issued forecasts for firm *j* in year *t*.[28] Other variables are identical to those in Eq. (1), but the analyst-level control variables are aggregated to the firm-year level by taking the average.

Table 3-5 presents the results of regressions of analyst following on corporate integrity culture and other control variables. The dependent variable is measured respectively at year t to year t+3 to examine the long-term impact. The results show significantly negative coefficients (at the 1% level) on *IntegrityCulture* for *Analyst Cov* in year t to year t+3 (in Columns 1 through 4). This suggests a negative relationship between integrity culture and analyst following, which remains for a long-term period (at least for next three years). This is mostly consistent with the argument that the demand for analysts' service is decreased for firms with a stronger integrity culture.

[28] As a robustness test, I calculate an alternative measure of analyst coverage: the number of earnings forecasts per month. The regression results using it as new dependent variables are consistent with those in Table 3-5, indicating that analysts issue fewer earnings forecasts for firms with a stronger integrity culture.

Table 3-5 Integrity Culture and Analyst Following

This table presents the results of regressions of analyst following on corporate integrity culture and other control variables. The dependent variables is *Analyst Cov*, calculated as the natural logarithm of one plus the number of unique analysts who have issued forecasts for the firm in a given year. The independent variable is *IntegrityCulture*, defined as the score of firm's integrity culture in a given year. The analyst-level controls include *Broker Size*, *Firm Experience*, *General Experience*, *Firms Followed*, *Industries Followed*, *Forecast Horizon*, *Days Elapsed*, *Forecast Frequency*, and *Lag AFE*. The analyst-level controls are aggregated at firm-year level by taking the average. The firm-level controls include *Size*, *Market-Book ratio*, *Growth, Tading Volume, Inst Own, ABR 12month, Restatement, Meet*, *Auditor Quality*, *CompAcctInd, Unexpected Earn, Neg UE, Neg SpItems*, *Loss*, *Volatility Earn*, *Volatility Ret*, *Bog Index*, *Document Length*, *Segments*, *Analyst Cov, Adv Exp, Miss_ Adv Exp, RD Exp* and *Miss_ RD Exp*. For definitions of these variables, please refer to Table C-1. All continuous variables are winsorized at the 1st and 99th percentiles. The firm and year fixed effects are added. For brevity, analyst-level and firm-level controls are not reported. The standard error is clustered at the firm level. T-statistics are reported in parentheses. *, **, and *** indicate statistical significance at the 10%, 5%, and 1% levels, respectively.

	(1)	(2)	(3)	(4)
	Analyst Cov	Analyst Cov $_{(t+1)}$	Analyst Cov $_{(t+2)}$	Analyst Cov $_{(t+3)}$
IntegrityCulture	-0.010***	-0.014***	-0.010***	-0.010***
	(-4.522)	(-5.504)	(-3.191)	(-2.968)
Analyst Controls	Yes	Yes	Yes	Yes
Firm Controls	Yes	Yes	Yes	Yes
Firm fixed effects	Yes	Yes	Yes	Yes
Year fixed effects	Yes	Yes	Yes	Yes
Number of observations	32,229	28,486	24,849	21,811
Adjusted R-squared	0.900	0.881	0.870	0.868

Alternatively, there might be a reverse relationship between corporate integrity culture and analyst following that the larger number of analysts covering the firm might have a negative impact on corporate integrity culture. Specifically, as discussed in the Section 2.3.5, analysts can put short-term pressure on firm executives, because firms' executives could be punished if they fail to meet analysts' earnings forecasts (He and Tian, 2013; Qian et al., 2019). Additionally, firm executives might be required to choose between long-term firm values and short-term profits, because they are not always compatible with each other. As a result, under the short-term pressure imposed by analysts, firm executives are more likely to choose the short-term projects that could provide immediate profits, instead of taking actions that might not generate instant benefits but are beneficial in the long-term period, such as investing more time or resources in constructing the corporate culture. Consistently, the results in the Section 2.5.2 indicate negative association between analyst coverage and corporate integrity culture (as well as other four cultural values), but such relationship is significant only in the current year (in year t) and becomes insignificant in the longer period.[29]

Taken together, the results in Table 3-5 are more consistent with my argument that analysts' service is less demanded for firms with a stronger integrity culture. According to the arguments discussed above, the lower demand of analysts' service might encourage analysts to make more efforts to collect and analyze new information, or to take anti-herding behavior.

3.7.3. Analyst forecast error

In this section, I examine the association between analysts' absolute forecast error and the covered firms' integrity culture. On the one hand, the better information environment of

[29] As a robustness test, I further conduct the regression of the change in analyst coverage (*Analyst Cov*) on the lagged change in the score of corporate integrity culture (*IntegrityCulture*) and lagged changes in other control variables. The results are consistent with the above results in Table 3-5.

firms with a stronger integrity culture is likely to reduce the information gap between analysts and these firms, which increases analysts' forecast accuracy (i.e., decreased absolute forecast error). However, if analysts covering these firms tend to have anti-herding behavior by strategically deviating their forecasts from others, they are likely to have less accurate forecasts, leading to higher absolute forecast error. The following OLS regression model is conducted to examine whether analysts have higher absolute forecast error when covering firms with a stronger integrity culture:

$$AFE_{i,j,t} = \alpha + \beta_1 IntegrityCulture_{j,t} + Controls + Fixed\ effects + \varepsilon_{i,j,t}. \quad (4)$$

The new dependent variable, AFE, is calculated as the absolute difference between an analyst's current forecast and the firm's actual EPS for the fiscal year-end, scaled by the firm's stock price two trading days before the forecast date. The AFE is multiplied by 100 to improve the readability of the coefficients of variables. Other variables are identical to those in Eq. (1).

Table 3-6 presents the results of regressions of analysts' absolute forecast error on corporate integrity culture and other control variables. The significantly positive coefficients on *IntegrityCulture* (0.060 and 0.059) in Columns 1 and 2 indicate that analyst forecasts are less accurate (with higher absolute forecast error) for firms with a stronger integrity culture, after controlling for analyst attributes and corporate characteristics. After including the analyst-year fixed effect, the coefficient on *IntegrityCulture* (0.029) remain positive but become insignificant, suggesting that the unobserved time-varying analyst characteristics could to some extent explain the association between analysts' absolute forecast error and the covered firms' integrity culture.

Table 3-6 Integrity Culture and Analyst Forecast Accuracy

This table presents the results of regressions of analyst forecast accuracy on corporate integrity culture and other control variables. The dependent variables is *AFE*, defined as the absolute difference between an analyst's forecast for the firm and the firm's actual earnings per share in the given year, scaled by the stock price two trading days before the forecast date. *AFE* is multiplied by 100 to improve the readability of the coefficients of variables. The independent variable is *IntegrityCulture*, defined as the score of firm's integrity culture in a given year. The analyst-level controls include *Broker Size*, *Firm Experience*, *General Experience*, *Firms Followed*, *Industries Followed*, *Forecast Horizon*, *Days Elapsed*, *Forecast Frequency*, and *Lag AFE*. The firm-level controls include *Size*, *Market-Book ratio*, *Growth, Tading Volume, Inst Own, ABR 12month, Restatement, Meet*, *Auditor Quality*, *CompAcctInd, Unexpected Earn, Neg UE, Neg SpItems*, *Loss*, *Volatility Earn*, *Volatility Ret*, *Bog Index*, *Analyst Cov, Document Length*, *Segments*, *Analyst Cov, Adv Exp, Miss_ Adv Exp, RD Exp* and *Miss_ RD Exp*. For definitions of these variables, please refer to Table C-1. All continuous variables are winsorized at the 1st and 99th percentiles. The firm, broker, year, analyst, and analyst-year fixed effects are added. For brevity, analyst-level and firm-level controls are not reported. The standard error is clustered at both the analyst and firm level. T-statistics are reported in parentheses. *, **, and *** indicate statistical significance at the 10%, 5%, and 1% levels, respectively.

	(1)	(2)	(3)
	AFE	AFE	AFE
IntegrityCulture	0.060**	0.059***	0.029
	(2.559)	(2.585)	(1.325)
Analyst Controls	Yes	Yes	Yes
Firm Controls	Yes	Yes	Yes
Firm fixed effects	Yes	Yes	Yes
Broker fixed effects	Yes	Yes	Yes
Analyst fixed effects	No	Yes	No
Year fixed effects	Yes	Yes	No
Analyst-Year fixed effects	No	No	Yes
Number of observations	999,908	999,483	997,842
Adjusted R-squared	0.45	0.464	0.526

Furthermore, consistent with the prior literature, analysts' absolute forecast error is highly correlated with the forecast horizon (i.e., the number of days between an analyst's forecast date for the firm and the firm's fiscal year-end date). Compared to the results in Table 3-3, the coefficients on *Forecast Horizon* in Table 3-6 have higher values and greater statistical significance. In unreported results, I only retain each analyst's most recent forecast for the firm in a year and replicate the regression. The results are very similar.

Taken with Table 3-3, these results show that analyst forecasts are bolder yet less accurate for firms with a stronger integrity culture. This is mostly consistent with the anti-herding behavior that analysts strategically deviate their forecasts from others when covering firms with a stronger integrity culture.

3.8. Endogeneity issue

By far, the findings in this study suggest that analysts tend to issue bolder forecasts for firms with a stronger integrity culture. The method in this research is less subject to potential endogeneity concerns.

Specifically, though it is not feasible to control for all the possible variables due to the lack of data availability or efficient proxies, this research uses a broader set of control variables than prior studies that focus on analysts' forecast boldness. These controls capture different analyst attributes and the covered firms' systematic differences that are found to be associated with analysts' forecast boldness. Furthermore, the model in this study includes more controls to mitigate the concern that the observed relationship between analysts' forecast boldness and corporate integrity culture might be driven by other forms of corporate information environment. Though there is no relatively direct empirical evidence previously about whether or not these new controls are correlated with analysts' forecast boldness, they are found by prior studies to capture different forms of companies' information environment and thus to be associated with analysts' other forecasting behavior, such as analysts' forecast accuracy.

Moreover, several useful fixed effects are added in the model. The firm and broker fixed effects are included to control for time-invariant firm and broker characteristics, while the year fixed effect is used to account for the time-varying effect. I further add the analyst fixed effect to control for time-invariant analyst characteristics, such as analysts' gender, their past experiences before work, social convention of analysts' country of origin. Some of them are found to be related to analysts' forecasting behavior (e.g., (Kumar, 2010; Cao et al., 2023)). More importantly, the analyst-year fixed effect is included to further account for the time-variant analyst attributes, such as analyst age, inherent ability, risk preference, income, star analyst, political preference, and overconfidence. By adding this fixed effect, these relevant analyst attributes, together with other analyst characteristics that could not be observed or quantified, are controlled. Taken together, the above control variables and different fixed effects could absorb much of analyst- and firm-level heterogeneities in analysts' forecasting behavior.

Furthermore, I mitigate potential endogeneity issues using the two-stage least squares (2SLS) instrumental variable method. An ideal instrumental variable should be correlated with the independent variable but uncorrelated with the residual term of the dependent variable (i.e., the omitted variables). It is challenging to identify an instrumental variable that is theoretically correlated with the corporate integrity culture but not correlated with analysts' forecast boldness. However, based on the prior literature, I introduce two instrumental variables to capture the variations in firms' integrity culture that are exogenous to analysts' forecasting behavior. Though not necessarily the perfect instrumental variables, the following two instrumental variables could, at least to some extent, satisfy the conditions of relevance and exclusion restriction.

I first adopt the average corporate integrity cultural values within an industry (*Avg IntegrityCulture*) as one of the instrumental variables. Such method based on the industrial average is frequently used in the prior literature to construct the instrumental variable (e.g., (Hanlon et al., 2003; Cassar et al., 2015; Dhaliwal et al., 2016; Ertugrul et al., 2017)). In related to the research area of corporate culture, Jiang et al. (2017) use the industrial average to construct their instrumental variable for firms' integrity culture. They find that firms with culture of higher integrity are associated with lower investment–cash flow sensitivity.

Specifically, this instrumental variable is calculated as the average value of the scores of integrity culture for all the other firms (excluding the given firm itself) within the industry of the firm in a given year.[30] This instrumental variable is likely to be associated with the firm's integrity culture. In particular, if the peer companies within an industry mostly establish a strong integrity culture, then the given firm is more likely to consider the integrity culture as a necessary corporate component and the firm can be motivated to construct a strong integrity culture, for the purpose of maintaining its competitive advantage. Hence, this instrumental variable could capture the relatively exogenous variation in firms' integrity culture. Since this instrumental variable pertains mostly to the external industry attributes rather than the given firm's own characteristics, it is less likely to be correlated with the individual firm's omitted factors that may influence analysts' forecast boldness.

The second instrumental variable is the CEO age (*CEO Age*). The previous studies find that CEOs become more ethical and more conservative as they age (Sundaram and Yermack, 2007; Lee et al., 2012; Serfling, 2014). In addition, there are empirical evidences that CEO characteristics could greatly influence the firms' culture (Van den Steen, 2010b; O'Reilly et al., 2014; Davidson et al., 2015). A recent research by Graham et al. (2022b) has conducted the survey on firm executives and find that the current CEOs are regarded by these participating respondents as the most influential individuals in constructing the firm's current culture. Thus, I presume that firms are likely to have a stronger integrity culture as their CEO age. Moreover, to the best of my knowledge, no empirical evidences have found

[30] An alternative instrumental variable is the average corporate integrity cultural values within the same region, which is calculated as the average value of the scores of integrity culture for all the other firms (excluding the firm itself) that are located within the same area (sharing the first three digits of a 5-digit zip code) of the firm in a given year. However, in the unreported results, this instrumental variable is found to have no significant correlation with the score of corporate integrity culture.

that the age of corporate CEOs could directly affect the forecast boldness of analysts who cover the firm. Plus, this research has added a broad number of control variables that might be associated with analyst forecast boldness. Hence, the *CEO Age* is less likely to be correlated with the omitted variables that are contained in the residual term of the dependent variable (*Forecast Boldness*), suggesting that this instrumental variable is likely to satisfy the condition of exclusion restriction.

The results for the 2SLS regression are reported in Table 3-7. In the first-stage of the 2SLS regression (Panel A of Table 3-7), I regress the score of corporate integrity culture (*IntegrityCulture*) respectively on the two instrumental variables: *CEO Age* and *Avg IntegrityCulture*. Other control variables are identical to those in Eq. (1), while the analyst-level control variables are aggregated to the firm-year level by taking the average. The coefficients on *Avg IntegrityCulture* (0.075) (in Column 1) and on *CEO Age* (0.008) (in Column 2) are significantly positive at the 1% level, which is consistent with the prediction that these two instrumental variables are significantly correlated with the score of corporate integrity culture.

Table 3-7 Two-Stage Least Squares Regression

This table presents the results of two-stage least squares (2SLS) regressions of analyst forecast boldness on corporate integrity culture and other control variables. Panel A reports the first-stage of the 2SLS regression of corporate integrity culture on two instrumental variables (*Avg IntegrityCulture* and *CEO Age*) and other control variables. The dependent variable is *IntegrityCulture*, defined as the score of firm's integrity culture in a given year. *Avg IntegrityCulture* is calculated as the average value of the scores of integrity culture for all the other firms (excluding the given firm itself) within the industry of the firm in a given year. *CEO Age* is defined as the age of the CEO from the firm in the given year. The analyst-level and firm-level controls are the same as in Table 3-3. The analyst-level controls are aggregated at firm-year level by taking the average. The firm and year fixed effects are added. For brevity, analyst-level and firm-level controls are not reported. The standard error is clustered at the firm level. Panel B reports the second-stage of the 2SLS regression of analyst forecast boldness on the predicted corporate integrity culture (*IntegrityCulture (Instru 1)* and *IntegrityCulture (Instru 2)*) from the first-stage estimation, and other control variables. The dependent variables is *Forecast Boldness*, defined as the average of the absolute difference between an analyst's current forecast and (1) the most recent consensus forecast and (2) the analyst's most recent previous forecast, which is set to zero if the current forecast is not higher or lower than both of the two benchmarks. The analyst-level and firm-level controls are the same as in Table 3-3. The firm, broker, year, analyst, and analyst-year fixed effects are added. For brevity, analyst-level and firm-level controls are not reported. The standard error is clustered at both the analyst and firm level. For definitions of these variables, please refer to Table C-1. All continuous variables are winsorized at the 1st and 99th percentiles. T-statistics are reported in parentheses. *, **, and *** indicate statistical significance at the 10%, 5%, and 1% levels, respectively.

Panel A: First-stage of the 2SLS regression

	(1) IntegrityCulture	(2) IntegrityCulture
Avg IntegrityCulture	0.075***	
	(5.121)	
CEO Age		0.008***
		(4.795)
Analyst Controls	Yes	Yes
Firm Controls	Yes	Yes
Firm fixed effects	Yes	Yes
Year fixed effects	Yes	Yes
Number of observations	31,088	22,205
Adjusted R-squared	0.539	0.543

Panel B: Second-stage of the 2SLS regression

	(1) Forecast Boldness	(2) Forecast Boldness
IntegrityCulture (Instru 1)	0.639***	
	(4.310)	
IntegrityCulture (Instru 2)		0.321**
		(2.433)
Analyst Controls	Yes	Yes
Firm Controls	Yes	Yes
Firm fixed effects	Yes	Yes
Broker fixed effects	Yes	Yes
Analyst-Year fixed effects	Yes	Yes
Number of observations	974,127	819,205
Adjusted R-squared	0.359	0.353

To conduct the second-stage estimation, I replicate the regression in Eq. (1) by replacing the independent variable of interest, *IntegrityCulture*, with *IntegrityCulture (Instru 1)* and *IntegrityCulture (Instru 2).* The results are shown in Panel B of Table 3-7. The *IntegrityCulture (Instru 1)* and *IntegrityCulture (Instru 2)* are the predicted scores of corporate integrity culture (*IntegrityCulture*) estimated from the first-stage estimation. The results in Column 1 and 2 indicate that the coefficients on *IntegrityCulture (Instru 1)* (0.639) and *IntegrityCulture (Instru 2)* (0.321) are significantly positive at the 5% and 1% level, after controlling for a vast number of analyst and firm attributes. This is consistent with the baseline results in Table 3-3. Overall, the 2SLS results are consistent with the main finding that analysts are more likely to have bolder forecasts for firms with a stronger integrity culture.

3.9. Alternative measure of integrity culture

As a robustness test, I employ an alternative measure of the score of corporate integrity culture based on the analyst report transcripts. Analysts are required to discover and collect the information about firms, through channels such as analyzing the company' financial statement and other disclosed documents, direct communication with corporate executives during calls, and implementing site visits. Analysts' opinions or feelings about the company could be reflected in the textual content of their analyst reports. Thus, similar to the earnings conference call transcripts, the textual information in analyst report transcripts may also give some insight into the firm's integrity culture, such as how transparent the firms are when disclosing their information. Graham et al. (2022b) also recommend using the analyst report as one potential source to construct a measure of corporate culture.

The analyst report transcripts are collected from the Investext through Thomson One database. The sample contains the S&P 500 companies from 2015 to 2020. It includes the companies that have been once listed on the S&P 500 Index at any point during the period of 2015 to 2020. For each analyst report, I exclude the unnecessary content that is not analysts' opinions or readable texts, including graphs, charts, tables, blank pages, the general introduction of company, broker information, the analyst's personal information, glossary, appendix, analyst disclosure and disclaimer. Furthermore, all words are transferred into lower case, Meanwhile, the plural nouns are converted to singular forms, and verbs to their present tense. And the characters which are not words (including numbers, punctuations and special characters) are removed. Finally, I exclude the firm names, the words related to dates and numbers (e.g., fifteen and January), and the stops words that are commonly used but have very little economic meanings, such as "a," "and," "is," and "you." The more detailed description of most of the above procedure is discussed in Chapter one.

Afterwards, I use the integrity culture-related word list from Li et al. (2021b) as the integrity culture dictionary. And a similar procedure is implemented to obtain the score of corporate integrity culture based on the analyst report transcripts. For each analyst report transcript, I calculate the weighted count of the number of words that appear in the integrity culture dictionary. Similarly, the weighting method is the tf-idf (term frequency–inverse document frequency) weight. As discussed before, according to this weighting scheme, the unique words that show up in a smaller percentage of documents are given higher weight than the commonly used words across all documents (e.g., finance, firm, and profit). The score of the corporate integrity culture is defined as the weight count of the number of integrity culture-related words, scaled by the weighted count of all the words in the transcript. The computed scores of corporate integrity culture from analyst report transcripts are then aggregated to the firm-year level by calculating the average. The final score is multiplied by 100 to increase the readability of the coefficients of variables.

I then replicate the regression in Eq. (1) by replacing the *IntegrityCulture* with *IntegrityCulture (alter)*, which is the calculated score of corporate integrity culture based on the analyst report transcripts as discussed above. Other variables are identical to those in Eq. (1). The firm, broker, and analyst-year fixed effects are added. Table 3-8 reports the results of the regression of analysts' forecast boldness on the alternative measure of corporate integrity culture and other control variables.

Table 3-8 Alternative measure of Integrity Culture

This table presents the results of regressions of analyst forecast boldness on alternative measure of corporate integrity culture and other control variables. The dependent variables is *Forecast Boldness*, defined as the average of the absolute difference between an analyst's current forecast and (1) the most recent consensus forecast and (2) the analyst's most recent previous forecast, which is set to zero if the current forecast is not higher or lower than both of the two benchmarks. The independent variable is *IntegrityCulture (alter)*, defined as the score of firm's integrity culture calculated based on analyst report transcripts. The analyst-level controls include *Broker Size*, *Firm Experience*, *General Experience*, *Firms Followed*, *Industries Followed*, *Forecast Horizon*, *Days Elapsed*, *Forecast Frequency*, and *Lag AFE*. The firm-level controls include *Size*, *Market-Book ratio*, *Growth, Tading Volume, Inst Own, ABR 12month, Restatement, Meet*, *Auditor Quality*, *CompAcctInd, Unexpected Earn, Neg UE, Neg SpItems*, *Loss*, *Volatility Earn*, *Volatility Ret*, *Bog Index*, *Analyst Cov, Document Length*, *Segments*, *Analyst Cov, Adv Exp, Miss_ Adv Exp, RD Exp* and *Miss_ RD Exp*. For definitions of these variables, please refer to Table C-1. All continuous variables are winsorized at the 1st and 99th percentiles. The firm, broker, and analyst-year fixed effects are added. For brevity, analyst-level and firm-level controls are not reported. The standard error is clustered at both the analyst and firm level. T-statistics are reported in parentheses. *, **, and *** indicate statistical significance at the 10%, 5%, and 1% levels, respectively.

	(1) Forecast Boldness
IntegrityCulture (alter)	0.054
	(1.327)
Analyst Controls	Yes
Firm Controls	Yes
Firm fixed effects	Yes
Broker fixed effects	Yes
Analyst-Year fixed effects	Yes
Number of observations	119,912
Adjusted R-squared	0.422

The coefficient on *IntegrityCulture (alter)* (0.054) is positive, which is consistent with the baseline result, but the coefficient is not statistically significant. This might be due to the relatively small sample of transcripts for calculating the alternative integrity score in that the collected analyst report transcripts only cover the S&P 500 companies from 2015 to 2020. Hence, after aggregating the integrity scores of transcripts into firm-year level, it is difficult to observe adequate variation in the aggregated integrity score across firms or across time.

3.10. Concluding remarks

In this research, I examine whether analysts' forecasting behavior is affected by the covered firm's information environment that is characterized by a strong integrity culture. The baseline results demonstrate that analysts are more likely to issue bolder earnings forecasts for the firms with higher scores of corporate integrity culture. To further understand the underlying explanation for the increased analyst forecast boldness for these firms, other results show that analysts' forecasts boldness is associated with weaker market reactions for firms with higher scores of corporate integrity culture than other firms, which suggests that analysts tend to provide relatively less informative bold opinions for firms with a stronger integrity culture. This is consistent with analysts' anti-herding behavior for firms with a stronger integrity culture.

Moreover, other results indicate a negative association between the scores of firms' corporate integrity culture and the number of analysts covering the firms, indicating lower demand for analysts' services for companies with a stronger integrity culture, which might lead analysts to have anti-herding behavior. Additionally, I find that analysts tend to have lower forecast accuracy for firms with higher scores of corporate integrity culture, but this negative relationship is not statistically significant after controlling for the unobserved time-varying analyst attributes.

To further mitigate potential endogeneity concerns, I introduce two instrumental variables: the average score of corporate integrity culture in an industry and the CEO age of the firm. The results of two-stage least squares (2SLS) regression are consistent with the baseline results that analysts tend to have bolder forecasts for firms with a stronger integrity culture. Finally, as a robustness test, I use the analyst report transcripts to develop an alternative measure of the score of corporate integrity culture. The results are consistent with the baseline results, but the observed association is not statistically significant, possibly because the limited sample of analyst report transcripts could lead to inadequate variation in the aggregated integrity score across firms or years.

Overall, this research studies corporate integrity culture as a fundamental aspect of firms' information environment and investigates its impact on analysts' forecasting behavior. The empirical findings are consistent with that analysts tend to issue bolder forecasts for firms with stronger integrity cultures, illustrating how the firm's information environment characterized by a strong integrity culture could affect the external stakeholders' perceptions and behaviors. Furthermore, the research provides novel evidence of analysts' anti-herding behavior when covering these firms, which not only deepens our understanding of the conditions under which analysts engage in anti-herding behavior, but also highlights the unintended consequences of such corporate transparency, as well as the importance of considering corporate culture as a critical role that influences both internal and external firm interactions.

A potential limitation of this study is its dependence upon a textual analysis-based measure of corporate integrity culture, which is derived from conference call transcripts. Though this measure could alleviate some issues in previous methods, it might not fully capture a firm's integrity culture due to the complex and abstract nature. Additionally, the sample is limited to U.S. public companies, which may restrict the generalizability of the findings. Moreover, although the study employs instrumental variables and fixed effects to address endogeneity issue, there may still be unobserved factors that influence both corporate integrity culture and analysts' forecasting behavior, which distort the results.

Future research could investigate other aspects of analysts’ forecasting behavior, such as forecast optimism, timeliness, and consistency across quantitative metrics. This would provide a more comprehensive understanding of how corporate integrity culture influences analysts’ roles as information intermediaries. In addition, future research could explore how corporate integrity culture affects other external stakeholders, such as media, auditors, and credit rating agencies. Finally, researchers could include the behavioral factors (e.g., risk preferences) that influence analysts’ decision-making, exploring how they interact with corporate integrity culture to shape analysts’ forecasting behavior.

3.11.Appendices

Table C-1 Variable Definitions

Variable	Description
Forecast Boldness	The average of the absolute deviation of an analyst's one year ahead earnings forecast for the firm from (1) the most recent consensus forecast for the same firm and (2) the analyst's most recent previous forecast for the firm, which is set to zero if the analyst's forecast is not higher or lower than both of the above two benchmarks, with this average absolute deviation scaled by stock price two trading days before the forecast date.
|ABR|	The absolute value of abnormal return staring from the forecast revision date, where the abnormal return is calculated as the buy-and-hold return of the stock minus the buy-and-hold return on the value-weighted market index for the same period.
Analyst Cov	The natural logarithm of one plus the number of unique analysts who have issued forecasts for the firm in a given year.
AFE	The absolute difference between an analyst's forecast for the firm and the firm's actual earnings per share in the given year, scaled by the stock price two trading days before the forecast date.
IntegrityCulture	The corporate integrity culture score from Li et al. (2021b), defined as the average score of corporate integrity culture obtained from the conference call transcripts of the firm in a given year, where the score of corporate integrity culture for each conference call transcript is calculated as the weighted count of the number of words related to culture of integrity (e.g., Integrity, Ethics, Accountability, Trust, Honesty, Responsibility), scaled by total number of words.
Strong IntegrityCulture	An indicator variable equal to one if the firm's *IntegrityCulture* is among the top quartile across all firms in a given year, and zero otherwise.
Avg IntegrityCulture	The average value of the scores of integrity culture for all the other firms (excluding the given firm itself) within the industry of the firm in a given year.
IntegrityCulture (alter)	The average score of corporate integrity culture calculated based on the analyst report transcripts for the firm in a given year, where the score of corporate integrity culture for each analyst report transcript is calculated as the weighted count of the number of words related to culture of integrity (e.g., Integrity, Ethics, Accountability, Trust, Honesty, Responsibility), scaled by the weighted count of all the words in the transcript.
Broker Size	The number of analysts employed by the brokerage, as captured by the number of unique analysts from the brokerage who have issued at least one earnings forecast in the I/B/E/S database in a given year.
Firm Experience	The number of years an analyst has issued earnings forecasts for the firm.
General Experience	The number of years an analyst has issued earnings forecasts for any firms in the I/B/E/S database.

Firms Followed	The number of companies an analyst has issued earnings forecasts for in a given year.
Industries Followed	The number of two-digit SICs an analyst has issued earnings forecasts for in a given year.
Forecast Horizon	The number of days between an analyst's earnings forecast date for the firm and the firm's fiscal year-end date.
Days Elapsed	The number of days between an analyst's forecast date and the date of the most recent prior forecast by any other analyst for the firm.
Forecast Frequency	The number of earnings forecasts issued by an analyst for the firm in a given year.
Lag AFE	The absolute forecast error of the last forecast issued by an analyst for the firm in the prior year, where the absolute forecast error is calculated as the absolute difference between an analyst's forecast for the firm and the firm's actual earnings per share in the given year, scaled by the stock price two trading days before the forecast date.
Size	The natural logarithm value of corporate assets in a given year.
Market-Book Ratio	The market value of corporate equity in a given year divided by the book value of corporate equity.
Growth	The annual growth rate of firm revenue from the prior year.
Trading Volume	The monthly average trading volume of the firm in a given year scaled by the number of shares outstanding.
Inst Own	The percentage of shares held by institutional investors for the firm in a given year.
ABR 12month	The twelve-month abnormal return prior to the earnings forecast date, where the abnormal returns is calculated as the buy-and-hold return of the stock minus the buy-and-hold return on the value-weighted market index for the same period.
Restatement	An indicator variable equal to one if the firm restates its annual or quarterly financial statements in a given year, and zero otherwise.
Meet	An indicator variable equal to one if the firm's actual earnings exactly meet or just beat the most preceding consensus forecast by one cent in a given year, and zero otherwise.
Auditor Quality	An indicator variable equal to one if the firm is audited by a Big N auditor in a given year, and zero otherwise.
Compacctind	The median value of the *Compacct* for the firm across all firms in the same two-digit SIC industry in a given year developed by De Franco et al. (2011), where the *Compacct* is the pairwise accounting comparability between the firm and any one of the other firms in the same industry, which is calculated as the absolute difference between the predicted value of the regression of the firm's earnings on its return and the predicted value of the firm's earnings using the new coefficients from the regression of another corresponding firm's earnings on its return.
Unexpected Earn	The absolute difference between actual corporate earnings and the earnings in prior year, scaled by the stock price at the end of the prior year.
Neg UE	An indicator variable equal to one if the firm's actual earnings is lower than the earnings in prior year, and zero otherwise.

Neg SpItems	The absolute value of the special item in a given year scaled by the total assets, which is set to if the special item has positive value.
Loss	An indicator variable equal to one if the firm's actual earnings in a given year is negative, and zero otherwise.
Volatility Earn	The standard deviation of the firm's previous 16 quarterly earnings scaled by the total assets.
Volatility Ret	The standard deviation of the firm's previous 48 monthly stock returns.
Bog Index	The measure of readability created by Editor Software's plain English software, StyleWriter, the formula of which is based on several textual attributes such as the sentence length, the English style problems, the word difficulty, and the ease of understanding the texts.
Document Length	The natural logarithm of one plus the average number of words in the Q&A sections of the conference call transcripts of the firm in a given year.
Segments	The natural logarithm of the number of reported business segments for the firm in a given year.
Adv Exp	The advertising expense of the firm in a given year scaled by the operating expense, which is set to zero if the firm does not disclose the advertising expense.
Miss_Adv Exp	An indicator variable equal to one if the firm's advertising expense is missing in a given year, and zero otherwise.
RD Exp	The research and development expense of the firm in a given year scaled by the operating expense, which is set to zero if the firm does not disclose the research and development expense.
Miss_RD Exp	An indicator variable equal to one if the firm's research and development expense is missing in a given year, and zero otherwise.
Prior |ABR|	The absolute value of the ten-day abnormal return ending the day before the forecast revision date, where the abnormal return is calculated as the buy-and-hold return of the stock minus the buy-and-hold return on the value-weighted market index for the same period.
Num Forecast	The number of one year ahead earnings forecasts issued on the forecast revision date.
CEO Age	The age of the CEO from the firm in a given year.

Conclusion

This thesis consists of three chapters that explore issues in related to the financial analysts. The first research in Chapter one concentrates on the determinants and impacts of analysts' tendency to provide similar textual information in their reports. In this chapter, I first examine four possible determinants of the analyst report similarity, which include analysts' herding behavior, analysts' poor ability to collect and provide new information, analysts' learning behavior, and the significant firm-related news. The regression results are mostly consistent with that the above four factors are likely the determinants of the similarity among analyst reports. Furthermore, I study the implications of the analyst report similarity for the market by examining whether this textual similarity could influence the informativeness of analyst reports. The results show that the analyst report similarity is negatively associated with the short-term investor reactions after the report, suggesting that market investors consider the analyst report as less informative if the textual information in the report is more similar to that in other analysts' prior reports. In the additional analysis, I find that analysts' quantitative outputs (e.g., earnings forecasts) become less informative when there is less new information in analyst reports to support these quantitative outputs, suggesting that the market investors may use the textual information in analyst reports to help understand the analysts' quantitative outputs. Moreover, my results show that the negative relationship between analyst report similarity and the investor reactions is stronger when the covered firm's managers have more incentives to withhold relevant corporate information. However, I find that this negative relationship is mitigated if analysts are from larger brokers. Finally, I conduct a robustness test by using the short-term abnormal trading volume to capture the informativeness of analyst reports. Consistently, I find a negative association between analyst report similarity and the investor reactions. Taken together this chapter makes contributions to the extant literature about the textual information in analyst reports, analysts' information intermediary roles in the market, and analysts' herding behavior.

The second research in Chapter two examines the impact of analysts on corporate governance in the context of corporate culture. First, I investigate the association between analyst coverage and the score of the covered firm's culture. The regression results show that the level of analyst coverage is negatively associated with the score of corporate culture, which is consistent with the pressure effect that analysts could impose short-term pressure on the firm, leading to a weak corporate culture. Consistently, further results show that analysts' possible negative impact is stronger for the long-term oriented cultural values than other cultural values. Furthermore, for the purpose of alleviating the potential endogeneity problems in baseline results, I use the two-stage least squares model based on an instrumental variable, the expected coverage. The results of this model are consistent with the baseline results that the analyst coverage is negatively associated with the corporate culture. In addition, I employ a quasi-natural experiment based on two exogenous shocks to the number of analysts covering the firm, the brokerage closures and mergers. The results of the Difference-in-Difference model also indicate that the analyst coverage has a negative impact on the covered firms' corporate culture. These results in the endogeneity tests suggest that the observed negative association in baseline results is likely causal. In the following additional tests, I examine whether this negative impact could vary with several analyst and firm characteristics. The results show that this negative impact is mitigated when the firms are covered by analysts with more experience, when the firms are more likely to reach analysts' earnings forecasts, and when the firms tend to have a better corporate governance. Overall, this chapter contributes to the literature about the corporate culture and the implications of analyst coverage for firms.

The third research in Chapter three investigates whether analysts' forecasting behavior is affected by the covered firm's information environment that is characterized by a strong integrity culture. Specifically, I examine whether the firms' integrity culture is associated with analyst forecast boldness. The baseline results indicate that analysts tend to generate bolder earnings forecasts for firms that have higher scores of integrity culture. I next conduct several further tests to understand the underlying explanation for analysts' boldness for firms with a strong integrity culture. I first find that the market regards analysts' bolder forecasts for these firms as less informative, indicating that analysts' bolder opinions for these firms likely reflect their tendency to strategically differentiate themselves from others (i.e., anti-herding behavior). Next, the results show that the firms with a stronger integrity culture are associated with lower number of analysts covering the firms. This could suggest that analysts' service is less demanded by investors for these firms. In addition, I find that analysts tend to generate less accurate earnings forecasts for firms with a stronger integrity culture. To deal with the potential endogeneity issues, I introduce the average score of corporate integrity culture within an industry and the CEO age of the firm as two instrumental variables to conduct the two-stage least squares model. Based on that, I find positive association between the firm's integrity culture and analyst forecast boldness, which is consistent with the baseline results. Finally, I use the analyst report transcripts to construct an alternative measure of the score of corporate integrity culture. Based on this new measure, I find a similar positive association consistent with the baseline results, but this relationship is not statistically significant, possibly due to the limited sample of analyst report transcripts that covers the S&P 500 companies from 2015 to 2020. According to these empirical findings in my research, this chapter contributes to the literature about analysts' forecasting behavior in related to the covered firm's information environment, and to the literature about the impact of corporate culture.

This thesis presents several significant contributions to the literature on financial analysts' practices and behaviors, shedding new light on their roles as information intermediaries, the effects of their activities on corporate governance, and the factors that shape their forecasting behavior.

In Chapter one, the research moves away from the widely examined quantitative outputs, such as earnings forecasts and stock recommendations, to explore the textual content of analyst reports. This shift provides novel insights into analysts' tendency to exhibit similar behavior. Based on a large sample of analyst report transcripts, the study illustrates that this tendency extends beyond quantitative metrics to textual communication in analyst reports. The findings suggest that textual similarity among reports decreases their informativeness, stressing the value of originality and uniqueness in textual content. Additionally, the research shows that market investors highly value analysts' ability to collect new information beyond both corporate disclosures and other analysts' reports. These insights deepen the understanding of analysts' role as information intermediaries and highlight the importance of textual analysis in studying analysts' behavior and its implications.

Chapter two adds to the ongoing debate about the role of financial analysts by investigating their influence on corporate culture. Although analysts are normally considered as reducing information asymmetry and serving as external monitors, this research demonstrates that the short-term pressure imposed by analysts can have negative impact on corporate culture. This study further complements prior studies that link analyst coverage to reduced innovation by demonstrating that such coverage can undermine the firm's cultural values that foster innovation. By illustrating how external pressure from analysts affects internal corporate culture, this study reveals the unintended negative consequences of analysts' activities.

In Chapter three, the study concentrates on corporate integrity culture as a key characteristic of firms' information environment and explores its impact on analysts' forecasting behavior. The findings indicate that analysts are more likely to issue bolder forecasts for firms with stronger integrity cultures, reflecting how a firm's information environment, shaped by integrity, can influence external stakeholders' perceptions and actions. Moreover, the study provides new evidence of analysts' anti-herding behavior when covering such firms, indicating that analysts can adopt more aggressive forecasting strategies in response to the reduced demand for their services caused by higher corporate transparency.

This thesis may have several limitations. First, the study is constrained by its sample selection. All three chapters focus on the U.S. markets, which may restrict the applicability of the findings to other regions outside the U.S. market. The study in Chapter one is limited to S&P 500 companies from 2015 to 2020 due to time constraints and the difficulty of accessing extensive reports, which may have unique characteristics that do not generalize to other contexts. Second, the research in Chapter two and three relies on the textual analysis-based measure of corporate culture, which is derived from conference call transcripts. While this method could mitigate some limitations of survey-based approaches and other textual analysis methods, it may not fully capture the depth and complexity of corporate culture. Additionally, this measure focus on five specific cultural values (innovation, integrity, quality, respect, and teamwork) that are prominent among S&P 500 companies, potentially overlooking other minor cultural values. Finally, although the study employs methodological techniques such as short-window event studies, instrumental variables, quasi-natural experiments, and fixed effects to address endogeneity concerns, there may still be unobserved factors influencing both the independent variables and the dependent variables, which could reduce the robustness of the findings.

Given the discussion above, this thesis suggests some possible avenues for future research. Future study could concentrate on exploring a broader range of textual attributes using advanced textual analysis methods, to deepen the understanding of how textual content influences investor decision-making and the informativeness of analyst reports. This topic could be extended to examine how the development of artificial intelligence technologies reshapes the textual content of analyst reports and its impact on market investors. Additionally, future studies should expand the analysis by including a broader sample of firms, time periods, and regions to enhance the generalizability of findings.

Furthermore, future research could investigate the role of external stakeholders, such as government regulators, media, and auditors, in shaping corporate culture. This would provide a more comprehensive understanding of how external agencies interact with internal corporate governance. In addition, researchers could explore alternative approaches to capturing corporate culture, such as applying more advanced textual analysis to firm-related materials (e.g., analyst reports).

Finally, future research could investigate other aspects of analysts' forecasting behavior, such as forecast optimism, timeliness, and consistency, to better understand how corporate integrity culture influences analysts' role as information intermediaries. By including behavioral factors (e.g., risk preferences) that influence analysts' decision-making, researchers could explore how these factors interact with corporate culture to shape analysts' forecasting behavior.

References

Abraham, S., & Bamber , M. (2017). The Q&A: Under surveillance. *Accounting, Organizations and Society, 58*, 15-31.

Adhikari, B. K. (2016). Causal effect of analyst following on corporate social responsibility. *Journal of Corporate Finance, 41*, 201-216.

Ammann, M., Oesch, D., & Schmid, M. M. (2013). Product Market Competition, Corporate Governance, and Firm Value: Evidence from the EU Area. *European Financial Management, 19*(3), 452-469.

Arya, A., Glover, J., Mittendorf, B., & Narayanamoorthy, G. (2005). Unintended consequences of regulating disclosures: The case of Regulation Fair Disclosure. *Journal of Accounting and Public Policy, 24*(3), 243-252.

Asquith, P., Mikhail, M. B., & Au, A. S. (2005). Information content of equity analyst reports. *Journal of Financial Economics, 75*(2), 245-282.

Audi, R., Loughran, T., & McDonald, B. (2015). Trust, but Verify: MD&A Language and the Role of Trust in Corporate Culture. *Journal of Business Ethics, 139*(3), 551-561.

Balthazard, P. A., Goldman, A., Cooke, R. A., & Potter, R. E. (2006). Dysfunctional culture, dysfunctional organization. *Journal of Managerial Psychology, 21*(8), 709-732.

Bamber, M., & Abraham, S. (2020). On the "Realities" of Investor-Manager Interactivity: Baudrillard, Hyperreality, and Management Q&A Sessions. *Contemporary Accounting Research, 37*(2), 1290-1325.

Bargeron, L., Lehn, K., & Smith, J. (2015). Employee–management trust and M&A activity. *Journal of Corporate Finance, 35*, 389-406.

Barron, O. E., Byard, D., & Yu, Y. (2008). Earnings Surprises that Motivate Analysts to Reduce Average Forecast Error. *The Accounting Review, 83*(2), 303–325.

Bartov, E., Givoly, D., & Hayn, C. (2002). The rewards to meeting or beating earnings expectations. *Journal of Accounting and Economics, 33*(2), 173-204.

Baruffaldi, S., Simeth, M., & Wehrheim, D. (2023). Asymmetric Information and R&D Disclosure: Evidence from Scientific Publications. *Management Science*.

Behn, B. K., Choi, J. H., & Kang, T. (2008). Audit Quality and Properties of Analyst Earnings Forecasts. *The Accounting Review, 83*(2), 327-349.

Bellstam, G., Bhagat, S., & Cookson, J. A. (2020). A Text-Based Analysis of Corporate Innovation. *Management Science*.

Bernhardt, D., Campello, M., & Kutsoati, E. (2006). Who herds? *Journal of Financial Economics, 80*(3), 657-675.

Bonsall, S. B., Leone, A. J., Miller, B. P., & Rennekamp, K. (2017). A plain English measure of financial reporting readability. *Journal of Accounting and Economics, 63*(2-3), 329-357.

Bozanic, Z., & Thevenot, M. (2015). Qualitative Disclosure and Changes in Sell-Side Financial Analysts' Information Environment. *Contemporary Accounting Research, 32*(4), 1595-1616.

Bradley, D., Gokkaya, S., & Liu, X. I. (2017). Before an Analyst Becomes an Analyst: Does Industry Experience Matter? *The Journal of Finance, 72*(2), 751-792.

Bradley, D., Mao, C. X., & Zhang, C. (2022). Does Analyst Coverage Affect Workplace Safety? *Management Science, 68*(5), 3464-3487.
Brown, L. D., Call, A. C., Clement, M. B., & Sharp, N. Y. (2016). The activities of buy-side analysts and the determinants of their stock recommendations. *Journal of Accounting and Economics, 62*(1), 139-156.
Brown, S. V., & Tucker, J. W. (2011). Large-Sample Evidence on Firms' Year-over-Year MD&A Modifications. *Journal of Accounting Research, 49*(2), 309-346.
Camerer, C., & Vepsalainen, A. (1988). The economic efficiency of corporate culture. *Strategic Management Journal, 9*(S1), 115-126.
Cao, Y., Hao, R., & Yang, Y. G. (2023). National culture and analysts' forecasting. *Review of Accounting Studies*.
Cassar, G., Ittner, C. D., & Cavalluzzo, K. S. (2015). Alternative information sources and information asymmetry reduction: Evidence from small business debt. *Journal of Accounting and Economics, 59*(2-3), 242-263.
Chang, H. S., Donohoe, M., & Sougiannis, T. (2016). Do analysts understand the economic and reporting complexities of derivatives? *Journal of Accounting and Economics, 61*(2-3), 584-604.
Chang, Y. C., Ljungqvist, A., & Tseng, K. (2022). Do Corporate Disclosures Constrain Strategic Analyst Behavior? *Review of Financial Studies, 36*(8), 3163-3212.
Chatman, J. A. (1991). Matching People and Organizations: Selection and Socialization in Public Accounting Firms. *Administrative science quarterly, 36*(3), 459-484.
Chatman, J. A., Caldwell, D. F., O'Reilly, C. A., & Doerr, B. (2014). Parsing organizational culture: How the norm for adaptability influences the relationship between culture consensus and financial performance in high-technology firms. *Journal of Organizational Behavior, 35*(6), 785-808.
Chatman, J. A., & Jehn, K. A. (1994). Assessing the Relationship between Industry Characteristics and Organizational Culture: How Different can You Be? *Academy of Management Journal, 37*(3), 522-553.
Chatman, J. A., & O'Reilly, C. A. (2016). Paradigm lost: Reinvigorating the study of organizational culture. *Research in Organizational Behavior, 36*, 199-224.
Chatman, J. A., Polzer, J. T., Barsade, S. G., & Neale, M. A. (1998). Being different yet feeling similar: The influence of demographic differences and organizational culture on work processes and outcomes. *Administrative science quarterly, 43*(4), 749-780.
Chen, N. X., Chiu, P.-C., & Shevlin, T. (2018). Do Analysts Matter for Corporate Tax Planning? Evidence from a Natural Experiment. *Contemporary Accounting Research, 35*(2), 794-829.
Chen, S., Matsumoto, D., & Rajgopal, S. (2011). Is silence golden? An empirical analysis of firms that stop giving quarterly earnings guidance. *Journal of Accounting and Economics, 51*(1-2), 134-150.
Chen, T., Harford, J., & Lin, C. (2015). Do analysts matter for governance? Evidence from natural experiments. *Journal of Financial Economics, 115*(2), 383-410.
Chen, X., Cheng, Q., & Lo, K. (2010). On the relationship between analyst reports and corporate disclosures: Exploring the roles of information discovery and interpretation. *Journal of Accounting and Economics, 49*(3), 206-226.
Christoffersen, J., & Stæhr, S. (2019). Individual risk tolerance and herding behaviors in

financial forecasts. *European Financial Management, 25*(5), 1348-1377.
Clarke, J., & Subramanian, A. (2006). Dynamic forecasting behavior by analysts: Theory and evidence. *Journal of Financial Economics, 80*(1), 81-113.
Cleary, S., Jona, J., Lee, G., & Shemesh, J. (2020). Underlying risk preferences and analyst risk‐taking behavior. *Journal of Business Finance and Accounting, 47*, 949–981.
Clement, M. B. (1999). Analyst forecast accuracy: Do ability, resources, and portfolio complexity matter? *Journal of Accounting & Economics, 27*(3), 285-303.
Clement, M. B., Hales, J., & Xue, Y. (2011). Understanding analysts' use of stock returns and other analysts' revisions when forecasting earnings. *Journal of Accounting and Economics, 51*(3), 279-299.
Clement, M. B., & Tse, S. Y. (2003). Do Investors Respond to Analysts' Forecast Revisions as If Forecast Accuracy Is All That Matters. *The Accounting Review, 78*(1), 227-249.
Clement, M. B., & Tse, S. Y. (2005). Financial analyst characteristics and herding behavior in forecasting. *Journal of Finance, 60*(1), 307-341.
Coleman, B., Merkley, K., & Pacelli, J. (2021). Human versus Machine A Comparison of Robo-Analyst and Traditional Research Analyst Investment Recommendations. *The Accounting Review*.
Cooke, R. A., & Lafferty, J. C. (1987). Organizational Culture Inventory (Form III). *Human Synergistics, Plymouth, MI.*
Cooper, R. A., Day, T. E., & Lewis, C. M. (2001). Following the leader: a study of individual analysts' earnings forecasts. *Journal of Financial Economics*.
Cuculiza, C., Antoniou, C., Kumar, A., & Maligkris, A. (2020). Terrorist Attacks, Analyst Sentiment, and Earnings Forecasts. *Management Science*.
Da, Z., & Huang, X. (2020). Harnessing the Wisdom of Crowds. *Management Science, 66*(5), 1847-1867.
Davidson, R., Dey, A., & Smith, A. (2015). Executives' "off-the-job" behavior, corporate culture, and financial reporting risk. *Journal of Financial Economics, 117*(1), 5-28.
De Franco, G., Hope, O.-K., Vyas, D., & Zhou, Y. (2015). Analyst report readability. *Contemporary Accounting Research, 32*(1), 76-104.
De Franco, G. U. S., Kothari, S. P., & Verdi, R. S. (2011). The Benefits of Financial Statement Comparability. *Journal of Accounting Research, 49*(4), 895-931.
Denison, D. R. (1984). Bringing corporate culture to the bottom line. *Organizational Dynamics, 13*(2), 5-22.
Denison, D. R., & Mishra, A. K. (1995). Toward a Theory of Organizational Culture and Effectiveness. *Organization Science, 6*(2), 204-223.
Derrien, F., & KecskÉS, A. (2013). The Real Effects of Financial Shocks: Evidence from Exogenous Changes in Analyst Coverage. *The Journal of Finance, 68*(4), 1407-1440.
Deshpandé, R., Farley, J. U., & Webster Jr, F. E. (1993). Corporate Culture, Customer Orientation, and Innovativeness in Japanese Firms: A Quadrad Analysis. *Journal of Marketing, 57*(1), 23-37.
Deshpandé, R., & Webster Jr, F. E. (1989). Organizational Culture and Marketing: Defining the Research Agenda. *Journal of Marketing, 53*(1), 3-15.
Dhaliwal, D., Judd, J. S., Serfling, M., & Shaikh, S. (2016). Customer concentration risk and the cost of equity capital. *Journal of Accounting and Economics, 61*(1), 23-48.

Dichev, I. D., & Tang, V. W. (2009). Earnings volatility and earnings predictability. *Journal of Accounting and Economics, 47*(1-2), 160-181.

Donelson, D. C., & Resutek, R. J. (2014). The predictive qualities of earnings volatility and earnings uncertainty. *Review of Accounting Studies, 20*(1), 470-500.

Doukas, J. A., & Zhang, R. (2021). Managerial ability, corporate social culture, and M&As. *Journal of Corporate Finance, 68*, 101942.

Edmans, A. (2011). Does the stock market fully value intangibles? Employee satisfaction and equity prices. *Journal of Financial Economics, 101*(3), 621-640.

Ellul, A., & Panayides, M. (2018). Do Financial Analysts Restrain Insiders' Informational Advantage? *Journal of Financial and Quantitative Analysis, 53*(1), 203-241.

Ertugrul, M., Lei, J., Qiu, J., & Wan, C. (2017). Annual Report Readability, Tone Ambiguity, and the Cost of Borrowing. *Journal of Financial and Quantitative Analysis, 52*(2), 811-836.

Fang, Y., Fiordelisi, F., Hasan, I., Leung, W. S., & Wong, G. (2023). Corporate culture and firm value: Evidence from crisis. *Journal of Banking & Finance, 146*, 106710.

Fich, E. M., Juergens, J. L., & Officer, M. S. (2018). Analyst Coverage and Acquisition Returns: Evidence from Natural Experiments. *Working paper*.

Fong, K., Hong, H., Kacperczyk, M., Kubik, J. D., & Ellul, A. (2022). Do Security Analysts Discipline Credit Rating Agencies? *The Review of Corporate Finance Studies, 11*(4), 815-848.

Frijns, B., & Huynh, T. D. (2018). Herding in analysts' recommendations: The role of media. *Journal of Banking and Finance, 91*, 1-18.

Garrett, J., Hoitash, R., & Prawitt, D. F. (2014). Trust and Financial Reporting Quality. *Journal of Accounting Research, 52*(5), 1087-1125.

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

Graham, J. R. (1999). Herding among investment newsletters: Theory and evidence. *Journal of Finance, 54*(1), 237-268.

Graham, J. R., Grennan, J., Harvey, C. R., & Rajgopal, S. (2022a). Corporate culture: Evidence from the field. *Journal of Financial Economics, 146*(2), 552-593.

Graham, J. R., Grennan, J. A., Harvey, C. R., & Rajgopal, S. (2022b). Corporate Culture: The Interview Evidence. *Journal of Applied Corporate Finance, 34*(4), 22-41.

Graham, J. R., & Harvey, C. R. (2001). The theory and practice of corporate finance: evidence from the field. *Journal of Financial Economics, 60*(2-3), 187-243.

Graham, J. R., Harvey, C. R., & Rajgopal, S. (2005). The economic implications of corporate financial reporting. *Journal of Accounting and Economics, 40*(1-3), 3-73.

Guiso, L., Sapienza, P., & Zingales, L. (2015a). Corporate Culture, Societal Culture, and Institutions. *American Economic Review, 105*(5), 336-339.

Guiso, L., Sapienza, P., & Zingales, L. (2015b). The value of corporate culture. *Journal of Financial Economics, 117*(1), 60-76.

Guo, B., Pérez-Castrillo, D., & Toldrà-Simats, A. (2019). Firms' innovation strategy under the shadow of analyst coverage. *Journal of Financial Economics, 131*(2), 456-483.

Hahn, T., & Song, M. (2013). Regulation Fair Disclosure and analysts' reliance on earnings announcements. *Journal of Accounting and Public Policy, 32*(1), 26-43.

Hallman, N., Howe, J. S., & Wang, W. (2022). Analyst coverage and syndicated lending.

Review of Accounting Studies.
Hanlon, M., Rajgopal, S., & Shevlin, T. (2003). Are executive stock options associated with future earnings? *Journal of Accounting and Economics, 36*(1-3), 3-43.
He, J., & Tian, X. (2013). The dark side of analyst coverage: The case of innovation. *Journal of Financial Economics, 109*(3), 856-878.
He, W., Jackson, A. B., & Li, C. K. (2020). Hierarchy and Performance of Analyst Teams. *European Accounting Review*.
Hirshleifer, D., Levi, Y., Lourie, B., & Teoh, S. H. (2019). Decision fatigue and heuristic analyst forecasts. *Journal of Financial Economics, 133*(1), 83-98.
Hirshleifer, D., Low, A., & Teoh, S. H. (2012). Are overconfident CEOs better innovators? *The Journal of Finance, 67*(4), 1457-1498.
Hirshleifer, D., & Teoh, S. H. (2003). Herd Behaviour and Cascading in Capital Markets a Review and Synthesis.
Hoberg, G., & Phillips, G. (2010). Product Market Synergies and Competition in Mergers and Acquisitions: A Text-Based Analysis. *Review of Financial Studies, 23*(10), 3773-3811.
Hoberg, G., & Phillips, G. (2016). Text-Based Network Industries and Endogenous Product Differentiation. *Journal of Political Economy*.
Hoi, C. K., Wu, Q., & Zhang, H. (2013). Is Corporate Social Responsibility (CSR) Associated with Tax Avoidance? Evidence from Irresponsible CSR Activities. *The Accounting Review, 88*(6), 2025-2059.
Hong, H., & Kacperczyk, M. (2010). Competition and Bias. *The Quarterly Journal of Economics, 125*(4), 1683–1725.
Hong, H., & Kubik, J. D. (2003). Analyzing the Analysts Career Concerns and Biased Earnings Forecasts. *The Journal of Finance, LVIII*.
Hong, H., Kubik, J. D., & Solomon, A. (2000). Security analysts' career concerns and herding of earnings forecasts. *RAND Journal of Economics, 31*(1), 121-144.
Hu, M., Jiang, C., John, K., & Ju, M. (2022). Fiduciary Duty of Loyalty and Corporate Culture. *Working paper*.
Huang, A. H., Lehavy, R., Zang, A. Y., & Zheng, R. (2018). Analyst information discovery and interpretation roles: A topic modeling approach. *Management Science, 64*(6), 2833-2855.
Huang, A. H., Zang, A. Y., & Zheng, R. (2014). Evidence on the information content of text in analyst reports. *The Accounting Review, 89*(6), 2151-2180.
Huang, R., Krishnan, M. M., Shon, J., & Zhou, P. (2017a). Who herds? Who doesn't? Estimates of analysts' herding propensity in forecasting earnings. *Contemporary Accounting Research, 34*(1), 374-399.
Huang, S. X., Pereira, R., & Wang, C. (2017b). Analyst Coverage and the Likelihood of Meeting or Beating Analyst Earnings Forecasts. *Contemporary Accounting Research, 34*(2), 871-899.
Hutton, I., Jiang, D., & Kumar, A. (2015). Political Values, Culture, and Corporate Litigation. *Management Science, 61*(12), 2905-2925.
Imam, S., & Spence, C. (2016). Context, not predictions: a field study of financial analysts. *Accounting, Auditing & Accountability Journal, 29*(2), 226-247.
Irani, R. M., & Oesch, D. (2013). Monitoring and corporate disclosure: Evidence from a

natural experiment. *Journal of Financial Economics, 109*(2), 398-418.
Irani, R. M., & Oesch, D. (2016). Analyst Coverage and Real Earnings Management: Quasi-Experimental Evidence. *Journal of Financial and Quantitative Analysis, 51*(2), 589-627.
Iselin, M., Park, M., & Van Buskirk, A. (2021). Seemingly inconsistent analyst revisions. *Journal of Accounting and Economics, 71*(1), 101339.
Islam, E., & Zein, J. (2020). Inventor CEOs. *Journal of Financial Economics, 135*(2), 505-527.
Ivković, Z., & Jegadeesh, N. (2004). The timing and value of forecast and recommendation revisions. *Journal of Financial Economics, 73*(3), 433-463.
Jegadeesh, N., & Kim, W. (2010). Do analysts herd? An analysis of recommendations and market reactions. *Review of Financial Studies, 23*(2), 901-937.
Jiang, C., John, K., & Ju, M. (2022). Shareholder Litigation and Corporate Culture. *Working paper*.
Jiang, D., Kumar, A., & Law, K. K. F. (2015). Political contributions and analyst behavior. *Review of Accounting Studies, 21*(1), 37-88.
Jiang, F., Kim, K. A., Ma, Y., Nofsinger, J. R., & Shi, B. (2017). Corporate Culture and Investment–Cash Flow Sensitivity. *Journal of Business Ethics, 154*(2), 425-439.
Jiao, Y. (2024). Managing decision fatigue: Evidence from analysts' earnings forecasts. *Journal of Accounting and Economics, 77*(1), 101615.
Jing, C., Keasey, K., Lim, I., & Xu, B. (2022). Analyst Coverage and Corporate Environmental Policies. *Journal of Financial and Quantitative Analysis*.
Jung, T., Scott, T., Davies, H. T. O., Bower, P., Whalley, D., McNally, R., et al. (2009). Instruments for Exploring Organizational Culture: A Review of the Literature. *Public Administration Review, 69*(6), 1087-1096.
Kadous, K., Mercer, M., & Thayer, J. (2009). Is There Safety in Numbers? The Effects of Forecast Accuracy and Forecast Boldness on Financial Analysts' Credibility with Investors. *Contemporary Accounting Research, 26*(3), 933-968.
Kasznik, R., & McNichols, M. F. (2002). Does Meeting Earnings Expectations Matter? Evidence from Analyst Forecast Revisions and Share Prices. *Journal of Accounting Research, 40*(3), 727-759.
Kelly, B., & Ljungqvist, A. (2012). Testing Asymmetric-Information Asset Pricing Models. *Review of Financial Studies, 25*(5), 1366-1413.
Kerr, J., & Slocum Jr, J. W. (2005). Managing corporate culture through reward systems. *Academy of Management Perspectives, 19*(4), 130-138.
Kim, J.-B., Lu, L. Y., & Yu, Y. (2019). Analyst Coverage and Expected Crash Risk: Evidence from Exogenous Changes in Analyst Coverage. *The Accounting Review, 94*(4), 345-364.
Kim, S. J. L., & Yu, K. (2004). Corporate culture and organizational performance. *Journal of Managerial Psychology, 19*(4), 340-359.
Kotter, J. P., & Heskett, J. L. (1992). Corporate culture and performance. *New York: Free Press*.
Kumar, A. (2010). Self-Selection and the Forecasting Abilities of Female Equity Analysts. *Journal of Accounting Research, 48*(2), 393-435.
Kumar, A. R., Ville; Xu, Rosy. (2022). Social learning and analyst behavior. *Journal of*

Financial Economics.
Lang, M., & Lundholm, R. (1993). Cross-sectional determinants of analyst ratings of corporate disclosures. *Journal of Accounting Research, 31*(2), 246–271.
Laster, D., Bennett, P., & Geoum, I. S. (1999). Rational Bias in Macroeconomic Forecasts. *The Quarterly Journal of Economics, 114*(1), 293-318.
Lauver, K. J., & Kristof-Brown, A. (2001). Distinguishing between Employees' Perceptions of Person–Job and Person–Organization Fit. *Journal of Vocational Behavior, 59*(3), 454-470.
Lee, C.-C., Rose-Green, E., & Huang, H.-W. (2012). CEO Age and Financial Reporting Quality. *Accounting Horizons, 26*(4), 725-740.
Leece, R. D., & White, T. P. (2017). The effects of firms' information environment on analysts' herding behavior. *Review of Quantitative Finance and Accounting, 48*(2), 503-525.
Lehavy, R., Li, F., & Merkley, K. (2011). The Effect of Annual Report Readability on Analyst Following and the Properties of Their Earnings Forecasts. *The Accounting Review, 86*(3), 1087-1115.
Li, K. (2020). Does Information Asymmetry Impede Market Efficiency? Evidence from Analyst Coverage. *Journal of Banking & Finance, 118*, 105856.
Li, K., Liu, X., Mai, F., & Zhang, T. (2021a). The Role of Corporate Culture in Bad Times: Evidence from the COVID-19 Pandemic. *Journal of Financial and Quantitative Analysis, 56*(7), 2545-2583.
Li, K., Mai, F., Shen, R., & Yan, X. (2021b). Measuring Corporate Culture Using Machine Learning. *Review of Financial Studies, 34*(7), 3265–3315.
Li, X., Lin, C., & Zhan, X. (2019). Does Change in the Information Environment Affect Financing Choices? *Management Science, 65*(12), 5676-5696.
Li, X., Wang, S. S., & Wang, X. (2017). Trust and stock price crash risk: Evidence from China. *Journal of Banking & Finance, 76*, 74-91.
Liu, X. (2016). Corruption culture and corporate misconduct. *Journal of Financial Economics, 122*(2), 307-327.
Livnat, J., & Zhang, Y. (2012). Information interpretation or information discovery: Which role of analysts do investors value more? *Review of Accounting Studies, 17*(3), 612-641.
Lobo, G. J., Song, M., & Stanford, M. (2012). Accruals quality and analyst coverage. *Journal of Banking & Finance, 36*(2), 497-508.
Loncan, T. (2022). Product Market Competition and FDI Decisions. *Journal of Financial and Quantitative Analysis*, 1-40.
Loughran, T., & Mcdonald, B. (2011). When Is a Liability Not a Liability: Textual Analysis, Dictionaries, and 10-Ks. *Journal of Finance, 66*(1), 35-65.
Loughran, T., & McDonald, B. (2015). The Use of Word Lists in Textual Analysis. *Journal of Behavioral Finance, 16*(1), 1-11.
Martens, T. I. M., & Sextroh, C. J. (2021). Analyst Coverage Overlaps and Interfirm Information Spillovers. *Journal of Accounting Research, 59*(4), 1425-1480.
Matsumoto, D., Pronk, M., & Roelofsen, E. (2011). What Makes Conference Calls Useful? The Information Content of Managers' Presentations and Analysts' Discussion

Sessions. *The Accounting Review, 86*(4), 1383-1414.
Matsunaga, S. R., & Park, C. W. (2001). The Effect of Missing a Quarterly Earnings Benchmark on the CEO's Annual Bonus. *The Accounting Review, 76*(3), 313-332.
Mikhail, M. B., Walther, B. R., & Willis, R. H. (1997). Do security analysts improve their performance with experience. *Journal of Accounting Research, 35*, 131–157.
Mikhail, M. B., Walther, B. R., & Willis, R. H. (2003). The effect of experience on security analyst underreaction. *Journal of Accounting and Economics, 35*(1), 101-116.
Miller, G. S. (2002). Earnings Performance and Discretionary Disclosure. *Journal of Accounting Research, 40*(1), 173-204.
Moshirian, F., Tian, X., Zhang, B., & Zhang, W. (2021). Stock market liberalization and innovation. *Journal of Financial Economics, 139*(3), 985-1014.
Nolte, I., Nolte, S., & Vasios, M. (2014). Sell-side analysts' career concerns during banking stresses. *Journal of Banking & Finance, 49*, 424-441.
O'Reilly, C. A. (1989). Corporations, Culture, and Commitment: Motivation and Social Control in Organizations. *California management review, 31*(1), 4.
O'Reilly, C. A., Caldwell, D. F., Chatman, J. A., & Doerr, B. (2014). The Promise and Problems of Organizational Culture. *Group & Organization Management, 39*(6), 595-625.
O'Reilly, C. A., & Chatman, J. (1996). Culture as social control: Corporations, cults, and commitment. *Research in Organizational Behavior, 18*, 157-200.
O'Reilly, C. A., Chatman, J., & Caldwell, D. F. (1991). People and Organizational Culture: A Profile Comparison Approach to Assessing Person-Organization Fit. *Academy of Management Journal, 34*(3), 487-516.
Olson, D. M., & Waguespack, D. M. (2020). Strategic behavior by market intermediaries. *Strategic Management Journal*.
Ottaviani, M., & Sorensen, P. (2006). The strategy of professional forecasting. *Journal of Financial Economics, 81*(2), 441-466.
Peterson, K., Schmardebeck, R., & Wilks, T. J. (2015). The Earnings Quality and Information Processing Effects of Accounting Consistency. *The Accounting Review, 90*(6), 2483-2514.
Qian, C., Lu, L. Y., & Yu, Y. (2019). Financial analyst coverage and corporate social performance: Evidence from natural experiments. *Strategic Management Journal, 40*(13), 2271-2286.
Quinn, P. J. (2018). Shifting corporate culture: executive stock ownership plan adoptions and incentives to meet or just beat analysts' expectations. *Review of Accounting Studies, 23*(2), 654-685.
Sathe, V. (1983). Implications of corporate culture: A manager's guide to action. *Organizational Dynamics, 12*(2), 5-23.
Scharfstein, D. S., & Stein, J. C. (1990). Herd behavior and investment. *American Economic Review, 80*(3), 465-479.
Schein, E. H. (2017). Organizational culture and leadership 5th ed. *San Francisco, CA: Jossey-Bass.*
Schwartz, H., & Davis, S. M. (1981). Matching corporate culture and business strategy. *Organizational Dynamics, 10*(1), 30-48.
Serfling, M. A. (2014). CEO age and the riskiness of corporate policies. *Journal of*

Corporate Finance, 25, 251-273.

Shroff, P. K., Venkataraman, R., & Xin, B. (2014). Timeliness of Analysts' Forecasts: The Information Content of Delayed Forecasts. *Contemporary Accounting Research, 31*(1), 202-229.

Soltes, E. (2014). Private Interaction Between Firm Management and Sell-Side Analysts. *Journal of Accounting Research, 52*(1), 245-272.

Sørensen, J. B. (2002). The Strength of Corporate Culture and the Reliability of Firm Performance. *Administrative science quarterly, 47*(1), 70-91.

Stickel, S. E. (1990). Predicting individual analyst earnings forecasts. *Journal of Accounting Research, 28*(2), 409-417.

Stickel, S. E. (1995). The Anatomy of the Performance of Buy and Sell Recommendations. *Financial Analysts Journal*.

Sundaram, R. K., & Yermack, D. L. (2007). Pay Me Later: Inside Debt and Its Role in Managerial Compensation. *The Journal of Finance, 62*(4), 1551-1588.

To, T. Y., Navone, M., & Wu, E. (2018). Analyst coverage and the quality of corporate investment decisions. *Journal of Corporate Finance, 51*, 164-181.

Trice, H. M. (1984). Studying Organizational Cultures Through Rites and Ceremonials. *Academy of Management Review, 9*(4), 653-669.

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

Twedt, B., & Rees, L. (2012). Reading between the lines: An empirical examination of qualitative attributes of financial analysts' reports. *Journal of Accounting and Public Policy, 31*(1), 1-21.

Van den Steen, E. (2010a). Culture Clash: The Costs and Benefits of Homogeneity. *Management Science, 56*(10), 1718-1738.

Van den Steen, E. (2010b). On the origin of shared beliefs (and corporate culture). *RAND Journal of Economics, 41*(4), 617-648.

Wang, K. (2020). Is the Tone of Risk Disclosures in MD&As Relevant to Debt Markets? Evidence from the Pricing of Credit Default Swaps *. *Contemporary Accounting Research, 38*(2), 1465-1501.

Wang, Z., Sun, L., & Wei, K. C. J. (2020). Does competition induce analyst effort? evidence from a natural experiment of broker mergers. *Journal of Banking & Finance*.

Welch, I. (2000). Herding among security analysts. *Journal of Financial Economics, 58*(3), 369-396.

Wen, Y., & Tikoo, S. (2020). Corporate strategy uniqueness, analyst herding, and financing constraints. *Strategic Organization*, 147612702092801.

Wu, J. S., & Zang, A. Y. (2009). What determine financial analysts' career outcomes during mergers? *Journal of Accounting and Economics, 47*(1-2), 59-86.

Ye, C., & Yu, L. H. (2017). The effect of restatements on analyst behavior. *Journal of Business Finance & Accounting, 44*(7-8), 986-1014.

Yu, F. (2008). Analyst coverage and earnings management. *Journal of Financial Economics, 88*(2), 245-271.

Yusoff, I., Chen, C., Lai, K., Naiker, V., & Wang, J. (2023). Foreign exchange exposure and analysts' earnings forecasts. *Journal of Banking & Finance, 146*, 106715.

Zaman, R. (2023). When corporate culture matters: The case of stakeholder violations. *The*

British Accounting Review, 56(1), 101-188.
Zhang, X. F. (2010). Information Uncertainty and Analyst Forecast Behavior*. *Contemporary Accounting Research, 23*(2), 565-590.
Zwiebel, J. (1995). Corporate Conservatism and Relative Compensation. *Journal of Political Economy, 103*, 1-25.