TWO ESSAYS ON THE EFFECTS OF INFORMATION DISCLOSURES ON INVESTOR BEHAVIORS

SHUQI WANG

PhD

The Hong Kong Polytechnic University

2024

The Hong Kong Polytechnic University

School of Accounting and Finance

Two Essays on the Effects of Information Disclosures on Investor Behaviors

Shuqi Wang

A thesis submitted in partial fulfilment of the requirements for the degree of Doctor of Philosophy

April 2024

CERTIFICATE OF ORIGINALITY

I hereby declare that this thesis is my own work and that, to the best of my knowledge and belief, it reproduces no material previously published or written, nor material that has been accepted for the award of any other degree or diploma, except where due acknowledgement has been made in the text.

_________________ (Signed)

___Wang Shuqi__________ (Name of student)

Abstract

This thesis includes two essays on the impact of information disclosure on investors' overconfidence and belief updating regarding the future stock price crash risk. The first essay examines how EDGAR implementation affects retail investors' overconfidence through reducing the acquisition costs of fundamental information costs with online disclosure. The second essay investigates whether implied stock price crash risk is affected within a short window around the management guidance.

In the first essay, we investigate whether information acquisition costs impact retail investors' overconfidence. Overconfidence is one of the most common behavioral biases among market participants in financial markets. Overconfident individuals tend to overestimate the precision of their knowledge and information. Models of financial markets with overconfident traders imply high trading volume, high volatility, and low price informativeness, explaining market anomalies with empirical and experimental evidence. However, these studies assume overconfidence. Using the implementation of the Electronic Data Gathering, Analysis, and Retrieval (EDGAR) system as an exogenous event, we find that overconfidence, measured by retail investors' trading activities and post-trade performance of stocks, is significantly reduced after firms join the EDGAR platform. The main results hold for both staggered and stacked difference-in-difference analyses. To further support the idea that the reduction in overconfidence is related to the decrease in information acquisition costs and increase in information set, we conduct a subsample analysis by dividing the dataset based on the level of information asymmetry for each firm. The reduction in overconfidence is greater for young firms and growth firms. These findings shed light on the effect of information sets on the overconfidence of retail investors.

In the second essay, we investigate whether managers' voluntary disclosures affect

investors' subjective stock price crash risk. Theoretical models indicate a predictive relationship between a firm's information environment and stock price crash risk. These models predict that there will be more stock price crashes if firms have less transparent firm-specific bad news. Using option-implied skewness as a measure of investors' subjective stock price crash risk, we find that investors' perception of future stock price crash risk decreases immediately after managers disclose bad news and increases after the announcement of good news. Further cross-sectional analysis demonstrates that this decrease in ex-ante stock price crash risk can be attributed to the fact that fewer negative news items are hidden by firms. The increase in ex-ante skewness can potentially be explained if investors suspect that managers strategically hide bad news and disclose good news for their own benefit. To understand the change in investors' perception of future tail risks, we demonstrate that the change in ex-ante stock price crash risk following earnings guidance disclosure provides additional predictive power for the realized skewness level in the following month.

In summary, these two essays shed light on the impact of information disclosure on investors' trading bias and belief updating regarding to future stock return distribution, providing new empirical evidence.

Acknowledgement

I would like to thank my Chief supervisor, Prof. K.C. John Wei, for his guidance and support during my PhD journey. I am also thankful to my co-supervisor, Dr. Gang Li, for his patience, encouragement and continuous help in my study. I thank my BoE chair Prof. Jay Cao and two external examiners Prof. Yangyang Chen and Dr Harry Wu for their valuable comments. I am grateful for all the comments and suggestions from Prof. JC Lin, Prof. Jeffrey Ng, Dr. Timothy Chue and Dr. Linti Zhang. I also deeply appreciate Terrance Odean and Li An for generously providing trading data.

I would also like to thank my classmates, Nanqi Li, Fengqin Chen, Lei Li, Jihye Yoo, Yani Li, Danlei Yu, Aurelius Aaron for their genuine help and warm company over the past five years. The moments of joy and caring we shared gave me the courage through the challenging times, especially in the final year. Your presence and friendship made this journey more meaningful and memorable.

Lastly, I thank my family for their unconditional love and support— special thanks to my mom, who lightened my life with her existence.

Table of Contents

Chapter 1 What Drives Retail Investors' Overconfidence? The Role of Information Acquisition Costs

1.1 Introduction

Overconfidence is one of the most common behavioral biases in financial markets and among market participants.[1] Overconfident people tend to overestimate the precision of their knowledge and information (Fischhoff, Slovic and Lichtenstein, 1977). Models of financial markets with overconfident traders imply high trading volume, high volatility, and low asset price informativeness, and help explain market anomalies, which are strongly supported by empirical and experimental evidence.[2] Overconfidence is assumed to be exogenous in these studies. However, what causes retail investors to become overconfident is not well understood. The objective of this study is to fill the gap in literature[3].

Overconfidence of retail investors could be affected by the information they use to estimate stock value and make investment decisions. As discussed by Kahneman (2011), "overconfidence is another manifestation of WYSIATI: when we estimate a quantity, we rely on information that comes to mind and construct a coherent story in which the estimate makes sense". At the same time, "you build the best possible story from the information available to you, and if it is a good

[1] De Bondt and Thaler (1995) state that "perhaps the most robust finding in the psychology of judgment is that people are overconfident."

[2] Overconfidence is an essential element of behavioral finance models to explain overreaction, such as that in Daniel, Hirshleifer and Subrahmanyam (1998) and Scheinkman and Xiong (2003). Daniel, Hirshleifer and Subrahmanyam (2001) state that equilibrium asset-pricing models in which traders are overconfident about their information help explain various market anomalies. Odean (1998) finds that trading volume and volatility increase and price informativeness decreases when price takers, insiders, or market makers are overconfident. Barber and Odean (2000) find overconfident individual traders continue to trade despite the fact that their poor performance. Glaser and Weber (2007) argue that overconfident investors trade more using the survey data. Experiemntal evidence suggests that investors are more likely perform worse in trading when they overestimate the precision of their signals (Biais, Hilton, Mazurier and Pouget, 2005).

[3] Age and gender are the most frequently discussed demographic characteristics of investors in relation to overconfidence. Other antecedents including investors' knowledge and experience, see review paper by (Singh, Malik and Jha, 2024). However, limited attention has been given to the external environment as a potential determinant of overconfidence.

story, you believe it. Paradoxically, it is easier to construct a coherent story when you know little, when there are fewer pieces to fit into the puzzle" (Kahneman, 2011). Therefore, when retail investors have more comprehensive data of firms' fundamentals, it may become more challenging for them to integrate all the information into a coherent pattern. Consequently, this increased difficulty in fitting everything they know into a cohesive narrative is expected to result in lower levels of overconfidence among retail investors.

In this paper, we empirically examine how overconfidence of retail investors responds to an increase in the information they possess. Specifically, we investigate whether a change in information acquisition costs affects the overconfidence of retail investors. The literature shows that even for public information, such as financial statements, acquisition and analysis costs are not negligible and can influence investors' behaviors and market outcomes.[4] A reduction in the costs for retail investors to obtain firms' financial data results in a larger information set available to them.

We use the 1993–1996 staggered implementation of the Electronic Data Gathering, Analysis, and Retrieval (EDGAR) system by the Securities and Exchange Commission (SEC) as an exogenous shock to information acquisition costs and examine its effect on overconfidence of retail investors. We choose the EDGAR implementation as an exogenous shock for three reasons. First, the implementation was mandatory and conducted by the SEC. Public companies in the United States were randomly assigned to different phases for transitioning to the new disclosure channel, which helps address concerns of endogeneity. Second, the EDGAR implementation significantly reduced information acquisition costs and modernized corporate disclosures.

[4] See Blankespoor, deHaan and Marinovic (2020) for a review of the studies on monitoring, acquiring, and analyzing firm disclosures.

Previously, companies had to send paper copies of their financial information to the SEC, which stored them for investors to review. This method was inefficient and posed a risk of lost files. However, with the implementation of EDGAR, companies can upload their financial statements online, allowing investors to freely access and download them if they have an internet connection. Unlike other recent updates to disclosure channels, which may have limited impacts on information dissemination due to the existence of numerous low-cost online channels, the EDGAR implementation revolutionized how investors obtain firm financial information. Third, anecdotal evidence suggests that individual investors use EDGAR platform to download information during EDGAR implementation period. The New York Times reported in 1994 that a personal investment club called "Investors Alliance" downloaded financial data from EDGAR for its members, eliminating the need for individual searches. Additionally, Gao and Huang (2020) highlight that approximately 31% of download requests made during the EDGAR implementation period were sent by retail investors. In summary, the choice of the EDGAR implementation as an exogenous shock is justified by its mandatory nature, the substantial reduction in information acquisition costs, and the evidence of retail investor usage through anecdotal sources and research.

We examine the effect of information acquisition costs on overconfidence using a database that contains more than 1.8 million transaction records from 77,037 unique investor accounts of a major U.S. discount brokerage house between January 1991 and December 1996, which covers the EDGAR implementation period.[5] To account for potential changes in the investor base following the event, we exclusively consider retail investors who had engaged in trading activities prior to the EDGAR implementation. We use trading volume, dollar volume, and trading frequency as measures of retail investors' trading activities because overconfident investors tend to trade

[5] We thank Terrance Odean and Li An for generously providing the trading data.

excessively (Daniel, Hirshleifer and Subrahmanyam, 1998; Odean, 1998; Gervais and Odean, 2001; Scheinkman and Xiong, 2003; Grinblatt and Han, 2005). In addition, we use the post-trade performance of stocks as a supplementary measure of overconfidence, following Kumar (2009). Overconfident investors are likely to make errors in their investment decisions because they tend to overestimate the precision of their estimation of the stock value and firm performance. Consequently, stocks tend to exhibit better performance following the sell transactions executed by overconfident investors compared to the purchases, which can be captured by the difference between the stock returns following their trades. We use the 30-day post-trade dollar return difference as a supplementary indicator of overconfidence.

In the baseline analysis, we examine the effect of the EDGAR implementation on overconfidence using a staggered difference-in-differences ("diff-in-diff" hereafter) analysis. The empirical results show that the reduction in information acquisition costs resulting from the EDGAR implementation leads to a significant decrease in retail investors' overconfidence. Specifically, we observe a decrease in trading volume, dollar volume, and trading frequency, by approximately 7.8%, 5.6%, and 7.7% of their standard deviations, respectively, following the EDGAR implementation. Moreover, we examine the 30-day post-trade dollar return difference between sell and buy trades as an additional measure of overconfidence. The results demonstrate that, on average, the difference in dollar returns following sell compared to buy trades within the same quarter decreases by \$31.13. This amount represents approximately 2.5 times the mean and 4.4% of the standard deviation. This decrease indicates a significant improvement in trading performance and reduction in overconfidence of retail investors after the EDGAR implementation.

Baker, Larcker and Wang (2022) highlight potential bias in staggered diff-in-diff analyses when earlier-treated units act as controls for later-treated units. To address this issue, we perform a stacked diff-in-diff analysis, following Cengiz, Dube, Lindner and Zipperer (2019). The goal of stacked diff-in-diff analysis is to generate event-specific 2×2 datasets, which are then stacked. In each event-specific dataset, the control units are the not-yet-treated observations. The stacked diff-in-diff analysis estimates an aggregated treatment effect that does not suffer from the bias caused by the potentially problematic control groups in the staggered diff-in-diff analysis. The baseline results continue to hold in our stacked diff-in-diff analysis.

To further support that the reduction in overconfidence is related to the reduction in information acquisition costs and increase in information set, we conduct a subsample analysis by dividing the full sample based on the level of information asymmetry for each firm. As retail investors have an information-access disadvantage relative to institutional investors, we expect that the impact of reduced information acquisition costs is more pronounced for retail investors' trades involving stocks with high information asymmetry. We split the sample equally based on firm age and, separately, based on book-to-market equity. Studies suggest that young firms and growth firms (proxied by the inverse measure of book-to-market ratios) are more likely than their counterparts to exhibit high information asymmetry (Zhang, 2006; Gao and Liang, 2013) .The analysis reveals that the reduction in overconfidence following the EDGAR implementation is primarily observed for young and growth firms. These findings align with the argument that the decrease in retail investor overconfidence is related to the decrease in information acquisition costs resulting from the EDGAR implementation. With decreased costs and improved access to fundamental information, retail investors may find it more difficult to generate coherent expectations regarding the stock value that is perceived to be highly precise, leading to a reduction in overconfidence.

The conclusion is supported by several robustness tests. First, we remove the first implementation phase from our sample and repeat the baseline analysis. According to SEC Release No. 33-6977, some firms volunteered to participate in the first implementation phase. Therefore, these firms may not be as randomized as those in the other nine phases, as is required for the staggered diff-in-diff analysis. After removing firms in the first phase, the results remain consistent, indicating a decrease in overconfidence following the EDGAR implementation. These results suggest that the impact of the EDGAR implementation on reducing overconfidence is not solely driven by the firms in the first phase, which strengthens the validity of findings.

Second, we conduct the pre-event parallel trend and falsification tests using the artificial implementation dates two years earlier and two years later, respectively, than the actual implementation dates. If the decrease in overconfidence is related to the decrease in the information acquisition cost from the EAGAR implementation, we should not observe significant decreases in behavioral bias if we use these pseudo-events. We find no significant decreases in overconfidence around these two sets of artificial EDGAR implementation dates, indicating that the main findings are robust to the pre-trend assumption and falsification test. Third, we examine each EDGAR implementation phase separately and find overconfidence decreases in all phases. This finding indicates that the decrease in overconfidence after the EDGAR implementation in the main analysis is not driven by a particular phase.

This paper contributes to the literature on the mechanisms of overconfidence. Early studies explore the impacts of overconfidence in terms of trading activities. Barber and Odean (2000) show that overconfidence can explain the high trading level and the resulting poor performance of individual investors. Barber and Odean (2015) later provide the supporting evidence that overconfidence can explain the increase in trading and reduction in performance of online investors. However, there is limited empirical evidence on the determinants of overconfident trading in terms of external information environment. Our findings, which suggest that decreased information acquisition costs reduce retail investors' overconfidence, provide empirical evidence of the relation between information and overconfidence as discussed by Kahneman (2011) in the context of investment. Our findings support the argument that with access to more fundamental information, retail investors decrease their overconfidence when making investment decisions, as they are less likely to construct highly coherent but false estimations of stock value.

Our study also relates to the literature on the effect of EDGAR implementation on the behaviors of market participants and market outcomes. Gao and Huang (2020) find that the EDGAR implementation significantly increases the efficiency of information production by sell-side analysts and retail investors, facilitating broader information dissemination.[6] Kim, Ivkovich and Muravyev (2021) present evidence of a causal relation between information costs and stock anomalies related to accounting information. They show that the average alphas of 125 accounting anomalies decrease significantly after companies disclose their financial information digitally through the EDGAR system. However, no such decrease occurs for non-accounting anomalies. Chang, Hsiao, Ljungqvist and Tseng (2022) focus on disagreement and find that the EDGAR implementation helps reduce disagreement around earnings announcement dates and reduces stock price crash risk. Goldstein, Yang and Zuo (2023) examine the economic impact of the EDGAR implementation as an introduction of modern information technology into financial markets. They argue that such a broader dissemination of information decreases the cost of capital and increases equity fundraising and investment but reduces managerial learning from stock prices.

The remainder of this chapter is organized as follows. Section 1.2 introduces the background of the EDGAR system and reviews the related literature. Section 1.3 discusses the data and

[6] Chang, Ljungqvist and Tseng (2023) show that after the EDGAR implementation, analysts significantly reduce coverage and issue less optimistic, more accurate, less bold, and less informative forecasts.

measures of the variables. Section 1.4 presents the main analysis of the overconfidence of retail investors. Section 1.5 describes the robustness analyses, and Section 1.6 concludes this chapter.

1.2 EDGAR system and retail investors

In this section, we briefly discuss the background of the EDGAR system, the association between the EDGAR system and information acquisition costs, and the design of the staggered implementation.

1.2.1 The EDGAR system and information acquisition of retail investors

In the early 1990s, the SEC wanted to use modern information technology to improve the efficiency of firms' information disclosure and did so by introducing the EDGAR system. This online system requires public firms to digitally disclose their financial statements for market participants to access and download for free, substantially reducing information acquisition costs.

Before the EDGAR system was introduced, both the disclosure and the acquisition of firms' financial statements through the SEC were inefficient. U.S. public firms submitted paper copies of their financial statements to the SEC by mail or by person. After being reviewed by the SEC, those paper copies were stored in public reference rooms in Washington D.C., New York, and Chicago. Due to the limited number of paper copies of these financial statements, usually only one or two per location, and restrictions on how many files could a person check out, only one company's documents could be inspected at a time. As a result, it was time-consuming or even impossible for investors, especially retail investors, to acquire the information in the financial statements in a timely manner.

Alternatively, retail investors could request paper copies of a company's financial statements by mail directly from the company. However, postal delivery can take a long time. In addition, it is not easy for investors to compare the financial statements of multiple public companies to make investment decisions.

The implementation of the EDGAR system revolutionized firms' information dissemination, bringing it into the digital age and significantly reducing information acquisition costs for investors, particularly retail investors. Through the EDGAR system, retail investors now have instant and free access to the financial information of all U.S. public firms. According to 58 F.R. 14628 (March 18, 1993, page 14,640), "Generally, as noted in the Proposing Release, public filings will be received, accepted and disseminated electronically on the same day." As further confirmed by the SEC's annual report, investors can obtain "10K/Q and all other corporate filings instantly on home computer screens" (Liu, 2019). Anecdotal evidence also suggests that retail investors use the EDGAR to acquire information when the system was launched. The New York Time (1994) reports "Investors Alliance, a personal-investment club in Fort Lauderdale, Florida, for example, downloads the 10 megabytes or so of new SEC material posted daily on the Internet and makes it available to users of its electronic bulletin board system. This saves individual members of the alliance from having to seek out the data themselves." Gao and Huang (2020) manually identify the domain names of retail investors associated with the searches of the filings in the EDGAR system during its implementation period and find that 24.45% of the total number of requests were made by retail investors, which accounts for 31.39% of the total amount of data requested.

1.2.2 The staggered EDGAR implementation

The EDGAR implementation occurred in ten phases, with the SEC randomly assigning all U.S. public firms to different phases. All firms in each phase were required to submit their filings to the SEC during the same period. After a pilot period in which some companies voluntarily disclosed their financial statements through EDGAR, the implementation process took three years, from 1993 to 1996. According to Appendix A of SEC Release No. 33-6977 (February 23, 1993), firms in the first phase (Group CF-01) had to meet the electronic filing requirements in April 1993, and firms in the last phase (Group CF-10) were required to file in May 1996. Starting in January 1994, thanks to the Internet Multicasting Service and New York University, retail investors can acquire the disclosures in EDGAR for free. We obtain the detailed EDGAR implementation phase-in schedule from Appendix B of SEC Release No. 33-6977.

Table 1.1 presents the timetable of the EDGAR system's implementation, including the implementation and effective time that we use in our empirical analysis. We follow the literature and conservatively set the effective time as two quarters after the implementation time in the SEC Release, on the assumption that investors must wait one quarter after a firm joins the EDGAR system for its most recent financial statements to become available online. For example, the implementation date for the first phase was April 26, 1993. This is the date that the firms in the first phase began filing their financial disclosures via EDGAR. At the beginning of the next quarter, 1993Q3, all firms in phase one should have completed the filing process to disclose their financial statements through EDGAR. Therefore, the earliest time that investors could access quarterly financial reports for 1993Q3 is 1993Q4.

[Insert Table 1.1 here]

1.3 Data and measures of the variables

In this section, we discuss the sample selection process. In addition, we describe measures of overconfidence and control variables.

1.3.1 Sample selection

To estimate how retail investors' trading behaviors are affected by the EDGAR implementation, we use the January 1991 to December 1996 trading data of 77,037 unique investor accounts with a major U.S. discount brokerage house. This dataset was introduced by Odean (1998) and later used by others to broadly examine retail investors' trading behaviors, such as overconfidence and disposition effect.[7] It includes records of all trades made through the discount brokerage house from 1991 to 1996. Each record contains an account identifier, the brokerage house's internal number for the security traded, a trade date, a buy-sell indicator, the quantity traded, the price, and the security's CUSIP number. Among the securities held and traded in the sample, we only select those that are common equities of U.S. corporations with a share class of 10 or 11 that are listed on the New York Exchange, American Stock Exchange, or Nasdaq Stock Market-National Market system with an exchange code of 1, 2, or 3.

When measuring trading behaviors that suggest behavioral biases, we follow the literature and discard potential short-sale transactions. We calculate the position of every security for each investor using the transaction data for the period after each trade. If investor B holds a negative position in stock A, we discard all of investor B's trades of stock A because of the ambiguity of negative positions. A negative cumulative share position means that an investor either opened that position before the start of the sample period and closed it during the sample period or short-sold the stock (Ben-David and Hirshleifer, 2012). The EDGAR implementation might have attracted sophisticated investors who started buying and selling stocks once they could acquire firms' fundamental information online, leading investor base change with EDGAR implementation. We address this concern by selecting the retail investors who first bought stocks before the EDGAR implementation to adjust the change in investor base during the sample period.

We obtain financial statement data from Compustat, stock price data from the Center for

[7] See, for instance, Odean (1999), Barber and Odean (2000), Ben-David and Hirshleifer (2012), An (2016), Gao and Huang (2020).

Research in Security Prices (CRSP), and analyst coverage data from the Institutional Brokers' Estimate System (I/B/E/S).

1.3.2 Measuring overconfidence

One major manifestation of overconfidence is the overestimation of the precision of information about the value of financial assets. Odean (1998) finds that overconfidence increases the expected trading volume and market depth but decreases the expected utility of overconfident traders. The models of Daniel, Hirshleifer and Subrahmanyam (1998), Odean (1998), Gervais and Odean (2001) and Scheinkman and Xiong (2003) show that as overconfidence increases, traders increasingly weigh their own signals more heavily than those of others when calculating their posterior beliefs. Therefore, their posterior beliefs are more dispersed, suggesting that overconfident investors trade more than other investors. Statman, Thorley and Vorkink (2006) empirically show that overconfidence explains high trading volume. In the international setting, Chui, Titman and Wei (2010) argue that investors from countries with more individualistic cultures tend to be more overconfident, and they provide empirical evidence that individualism is associated with trading volume.

We therefore follow these studies and use three variables related to trading activities of retail investors to measure overconfidence. The first is share trading volume (*VO*), which represents the number of shares traded by retail investors in our sample, calculated as the natural logarithm of one plus the sum of shares traded (in thousands) in a quarter for the stocks that are traded at least once by the retail investors in our sample. The second is dollar volume (*DVO*), which is calculated as the natural logarithm of one plus the sum of dollar volume traded (in thousand dollars) in a quarter. The third is trading frequency (*Freq*), which is the natural logarithm of one plus the sum of number of trades for each stock in a quarter. If a stock is not traded by any retail investors in a quarter, the three trading-related measures are zero. We aggregate each of these three measures at the stock-quarter level.

Investors who are overconfident about the quality of information or their ability to process it tend to make systematic mistakes, as discussed by Odean (1999). It is likely that the stocks sold by overconfident investors will systematically outperform the stocks they purchase in a significant manner. Following this idea, Kumar (2009) measures the investors' overconfidence by measuring the difference between the mean return following the sell trades and the mean return following the purchase trades at the end of each year. This study finds that the performance difference tends to be larger when there is higher valuation uncertainty in stocks, indicating a greater degree of overconfidence.

To capture the overconfidence of retail investors, we also estimate the post-trade performance of stocks following Kumar (2009). We calculate the 30-day post-trade sell-buy return differential (*PTSBD*) as the proxy for overconfidence as follows,

$$PTSBD_{i,t} = \frac{\sum_{j=1}^{N} PTSBD_{j,i,t}}{N}, \tag{1.1}$$

where

$$\begin{aligned} PTSBD_{j,i,t} = &\sum_{d=1}^{t} Price_{j,i,d}^{Sell} \times Shr_{j,i,d}^{Sell} \times Ret_{i,d+1,d+30} \\ &- \sum_{d=1}^{t} Price_{j,i,d}^{Purchase} \times Shr_{j,i,d}^{Purchase} \times Ret_{i,d+1,d+30}. \end{aligned} \tag{1.2}$$

Here $PTSBD_{j,i,t}$ is the *PTSBD* for stock i traded by retail investor j in quarter t, whereas $PTSBD_{i,t}$ represents the average *PTSBD* for stock i in quarter t. For each stock, we first calculate the 30-day returns following sells and purchases separately from retail investors in our sample following Eq. (1.2). $Price_{j,i,d}^{Purchase}$ or $Price_{j,i,d}^{Sell}$ is the actual purchase or sell prices of stock i made by investor

j on day d. If there are multiple purchases or sells of stock i by investor j on day d, we use the average price of purchases or sells. $Shr_{j,i,d}^{Purchase}$ or $Shr_{j,i,d}^{Sell}$ is the number of shares purchased or sold by investor j on stock i on day d. If there are multiple trades for stock i by investor j on day d, we take the sum of the number of shares. $Ret_{i,d+1,d+30}$ is the cumulative excess return of stock i in 30 days following the trade. With the performance measure estimated at the investor-stock level, we then aggregate the performance at the stock level in each quarter following Eq. (1.1) and winsorize it at the 1 and 99 percentile levels to eliminate the potential concerns from extreme values.

1.3.3 Control variables

We include several control variables in the regression models. As the EDGAR system is implemented approximately every quarter, the dependent variables are computed at the quarter level for each stock. For the control variables that are computed monthly for individual stocks, we use the last available monthly data for the quarter.

To control the impact of systematic risk on investor trading behaviors, we include the market beta (*Beta*), firm size (*Size*), book-to-market equity (*B/M*), and momentum (*MOM*). Following Fama and French (1992), we estimate *Beta* for each stock using its monthly returns over the previous 60 months. *Size* is computed as the natural logarithm of the market value of equity, calculated in million dollars. *B/M* is the natural logarithm of a firm's book equity at the end of the previous fiscal year, divided by the market value of equity at the end of December of the previous year.

We also control for the valuation uncertainty of stocks, as measured by trading volume turnover (*Turnover*) and stock idiosyncratic volatility (*IVOL*), following Kumar (2009), who argues that the behavior bias is stronger when valuation uncertainty is greater. *Turnover* is measured as the daily average number of shares traded in a quarter divided by the number of shares outstanding, expressed as a percentage. We divide the number of shares traded by 2 for Nasdaq stocks to address the double counting issue following Gao and Ritter (2010). Following Ang, Hodrick, Xing and Zhang (2006), monthly stock idiosyncratic volatility is computed as the standard deviation of the daily residuals in a month from the Fama and French (1993) three-factor model. As suggested by Kumar (2009), we include control variables that may affect investors' overconfidence. Firms listed on the Nasdaq stock exchange (*Nasdaq dummy* = 1) and those not paying dividends (*Dividend dummy* = 0) may be more growth-oriented, which may affect investors' overconfidence and trading behavior. Therefore, we include indicators for the stock exchange in the previous month and whether a firm pays dividends in the previous fiscal year. Analyst coverage (*ACov*) calculated as the natural logarithm of one plus the average number of analysts covering a stock in each quarter, and institutional holdings (*INST*) measured as the natural logarithm of one plus the institutional ownership scaled by the outstanding shares in each quarter, are also controlled because they are related to the information environments and may affect behavioral biases. Bid-ask spread (*Bid-ask spread*), calculated as the average daily closing bid-ask spread in each month for each stock in percentage, is included to control for the potential relation between microstructure effects and behavioral bias. Daily bid–ask spreads in the highest 1% tail of the distribution are eliminated for potential measurement errors.

Summary statistics of control variables and variables of interest are presented in Table 1.2 The mean trading volume (*VO*) is 0.84, that is, approximately 1,300 shares per quarter, which is small compared to the total trading volume of stocks because we only capture the number of shares traded by retail investors in our sample. The mean dollar volume (*DVO*) is 2.09 which is around

7,000 dollars trade by retail investors in a quarter for each stock. On average, each stock is traded 1.16 times in a quarter by retail investors. The mean of post-trade sell-buy return differential (*PTSBD*) is -12.3, indicating that the dollar amount earned in 30 days following the sells is 12.3 dollars less than the dollar amount earned in 30 days following the purchases. The median of the Post dummy variable is 0, and its mean is 0.36 because the EDGAR implementation started two years after the beginning of our sample period and occurred in phases. The mean *Beta* of the stocks in the sample is 1.07. The natural logarithm of the market value of the stocks in million dollars (*Size*) is 5.01. The stocks traded by retail investors in our sample have an average market value slightly higher than that of the common shares listed on the three major exchanges (*Size* ≈ 4.5) during the same period. The mean and median percentage returns for the preceding 11 months (*MOM*) are 9% and 10%, respectively. The mean of the natural logarithm of the book-to-market-equity ratios of the firms in our sample is -0.58, suggesting that on average the book value is less than the market value of equity. The mean and median price of the stock per share are 19.12 and 14.5, respectively.

[Insert Table 1.2 here]

Of the stocks in our sample, 51% pay dividends and 55% are listed on the Nasdaq stock exchange. The mean turnover is 0.27, indicating that on average, 0.27% of the outstanding shares of the firms in our sample change hands in a day. The mean of institutional ownership is 0.09, showing that the average percentage of shares held by institutional investors is 9%. The distribution of institutional ownership is positively skewed. The mean of analyst coverage is 0.84, suggesting that the average number of analysts following the stocks in our sample is 1.32. The average daily bid–ask spread is 4.26%. The average idiosyncratic volatility of the stocks in the sample is 2.93% per day. The average firm age (*Age*) measured as the number of years since the firm was first covered by the CRSP, is 15.98 years. In the following analysis, we use firm age to measure the degree of information asymmetry.

1.4 Empirical analysis

The EDGAR system provides an online platform for investors to access and download company financial information, thereby decreasing information acquisition costs. In this section, we use the staggered EDGAR implementation and transaction data of retail investors to examine the effect of the EDGAR implementation on retail investors' overconfidence.

1.4.1 Main results

In this section, we examine how the EDGAR implementation affects retail investors' overconfidence. EDGAR implementation reduces the acquisition costs of fundamental information for the general public, including retail investors. Through direct access to EDGAR over the Internet, individuals can obtain timely financial data. Kahneman (2011) states that overconfidence is determined by the coherence of the story one has constructed and it is easier to construct a coherent story when one has limited information. Therefore, we expect that overconfidence will decrease as retail investors gain more information about a firm's fundamentals. With more available information, the coherence of investors' expectations is likely to decrease.

To examine how the EDGAR implementation affects the overconfidence, we employ a staggered diff-in-diff analysis using the following OLS regression specification:

$$Y_{i,t} = \beta_1 \times Post_{i,t} + \beta_2 \times X_{i,t-1} + \gamma_i + \rho_t + \epsilon_{i,t}, \tag{1.3}$$

where $Y_{i,t}$ captures aggregate trading activities of stock i in quarter t. We use three measures for $Y_{i,t}$, trading volume (VO), dollar volume (DVO) and trading frequency ($Freq$). A decrease in each of the three measures indicates a reduction in overconfidence. $Post$ is a dummy variable that equals 1 if quarter t is in or after the effective year-quarter of the EDGAR implementation for firm i and zero otherwise. X_i represents a set of control variables. We include firm fixed effects (γ_i) to control for time-invariant differences in overconfidence across firms. We also include year-quarter fixed effects (ρ_t) to control for the effect of economy-wide shocks on overconfidence. To address time-series and cross-sectional correlation in the residuals, we cluster the standard errors by stock and year-quarter.

Table 1.3 presents the regression results of Eq. (1.3). We find consistent regression results across all three measures of trading activities. Columns (1) and (2) show the coefficients and corresponding t-statistics when the dependent variable is the trading volume (VO). The coefficient on $Post$ is -0.087 with a t-statistics of -4.35 in column (1). After controlling the relevant variables that could affect the overconfidence following previous literature, the coefficient, as shown in column (2) becomes -0.097 (t-stat =-5.18). In the economic terms, the coefficients on $Post$ in columns (1) and (2) indicate a decrease in trading volume by 7.76% and 8.66% of its standard deviation, respectively, after firms join the EDGAR system, without and with control variables considered. Columns (3) and (4) show the coefficients and corresponding t-statistics when the dependent variable is the dollar volume (DVO). The coefficient on $Post$ is -0.125 and -0.149, with t-statistics of -3.76 and -4.87, respectively. These coefficients show a decrease in DVO by 5.53% and 6.59% of its standard deviation after the EDGAR implementation. Similarly, we observe a decrease in trading frequency ($Freq$) by 7.65% and 8.71% of its standard deviation (columns 5 and 6), which is statistically significant at the 1% level. These negative coefficients indicate a statistically and economically significant reduction in overconfidence, as measured by trading activities, after firms join the EDGAR system and their financial information becomes available online.

[Insert Table 1.3 here]

Recent studies suggest that the empirical results of our staggered diff-in-diff analysis could be subject to bias due to problematic control groups (Baker, Larcker and Wang, 2022). Specifically, the coefficient on the dummy variable $Post$ is the variance-weighted average of the estimates for each implementation phase from the OLS regression. One concern highlighted in the literature is that the treatment effect observed in earlier implementation phases may not be solely indicative of the overall treatment effect. Instead, it may reflect changes in the effect over time. This issue raises the possibility that the coefficient on $Post$ in Eq. (1.3) captures differences in the treatment effect among different implementation phases rather than solely representing the overall treatment effect.

To address the potential concern regarding changes in treatment effects over time, one approach is to employ a stacked regression, as suggested by Cengiz, Dube, Lindner and Zipperer (2019) and Baker, Larcker and Wang (2022). The goal of a stacked diff-in-diff regression is to create a "clean" 2×2 dataset for each implementation phase. In each dataset, the treatment firms are those that are treated in a certain phase, while the control firms are not yet treated at the end of the relevant window (Cengiz, Dube, Lindner and Zipperer, 2019) al 2019). For each EDGAR implementation phase, we define the treated firms in the event-quarter window [-5, 5], where quarter 0 represents the EDGAR implementation quarter. Take the first implementation phase as an example. Since phase one's effective treatment date is 1993Q4, the window begins in 1992Q3 and ends in 1995Q1. In this case, the control group consists of firms in phases 6 through 10. The not-yet-treated firms in the first window are those in phases 6-10, as the earliest effective treatment date for phase 6 (1995Q2) occurs after the end of the window (1995Q1). Consequently, the treatment firms are those in phase one. To select the control firms, we employ the nearest-neighbor propensity score matching method, matching equity market capitalization (in levels and logs) by quarter, following Chang, Hsiao, Ljungqvist and Tseng (2022). We then generate the clean 2×2 event sample for the first six implementation phases during the 10 quarter-event windows around the effective date of each phase. However, since the transaction dataset ends in the treatment quarter of the last implementation phase, the last four implementation phases lack not-yet-treated control firms. Consequently, after stacking the dataset for each phase, our stacked dataset includes treatment effects for only the first six implementation phases.

We modify Eq. (1.3) and run the following regression on this stacked dataset:

$$Y_{i,t} = \beta_1 \times Post_{i,t} + \beta_2 \times X_{i,t-1} + \gamma_{i,p} + \rho_{t,p} + \epsilon_{i,t}, \tag{1.4}$$

where $\gamma_{i,p}$ and $\rho_{t,p}$ denote the firm and year-quarter fixed effects for implementation phase (p), which are defined as the interactions between the phase and firm dummies and between the phase and year-quarter dummies, respectively.

Table 1.4 presents the regression results for Eq. (1.4) using the stacked diff-in-diff analysis. The analysis results consistently confirm the findings obtained from the staggered diff-in-diff regressions. The dependent variables are trading volume, dollar volume, and trading frequency. Columns (1) and (2) of Table 1.4 show the coefficients and corresponding t-statistics when the dependent variable is trading volume (VO). The coefficient on $Post$ is -0.144 with the t-statistic of -4.60 in column (1). After controlling the variables that could affect the overconfidence following previous literature, the coefficient, as shown in column (2), becomes -0.138 (t-stat = -4.80). The coefficients on $Post$ without and with control variables indicate an average decrease in trading volume by 12.9% and 12.3% of its standard deviations after firms join the EDGAR system, respectively. Columns (3) and (4) show the coefficients and t-statistics when the dependent variable is dollar volume (DVO). The coefficient on $Post$ is -0.192 and -0.206 with t-statistics of -3.49 and -4.38, respectively, indicating on average dollar volume decreases by around 8.5% and

9.1% of its the standard deviation after the EDGAR implementation. Furthermore, trading frequency significantly decreases by 9.9% and 10.2% of its standard deviation as shown in columns (5) and (6), respectively. Overall, these results from the stacked diff-in-diff analysis are consistent with the findings from the staggered diff-in-diff regressions, confirming the robustness of our results. In conclusion, we find that the implementation of EDGAR helps decrease overconfidence among retail investors[8].

[Insert Table 1.4 here]

1.4.2 Heterogeneity in the effects of the EDGAR implementation

To better understand how the implementation of EDGAR reduces retail investors' overconfidence, we revisit our main analysis using different subsamples. As the EDGAR implementation decreases information acquisition costs and enhances the availability of firms' financial data, the decrease in trading activities after a firm's inclusion in EDGAR may be attributed to the increased availability of fundamental information about companies, which retail investors can use in forming their estimation of stock performance. With more financial data, retail investors may find it more challenging to construct a coherent narrative that aligns with their estimation using available fundamental information. This increased availability of information may, in turn, reduce overconfidence when making investment decisions. Therefore, we expect that the effect of the EDGAR implementation on overconfidence is stronger for firms with higher information asymmetry.

We use firm age (*Age*) and book-to-market equity (*B/M*) as proxies for information

[8] As institutional investors have access to the EDGAR information, retail investors may see themselves as disadvantaged in obtaining and analyzing the more comprehensive information available to sophisticated institutional investors. This perception could potentially reduce their overconfidence. In untabulated results, we find significant results that the overconfidence of retail investors decreases more when stocks are held by a larger number of institutional investors, presenting a potential solution to this concern.

asymmetry. Barry and Brown (1985) argue that the longer a firm has been listed on the stock market, the more information is available about it. Firm age, which is defined as the number of years since the firm was first covered by the CRSP, is used as an inverse proxy for information asymmetry in the literature (Leary and Roberts, 2010; Maskara and Mullineaux, 2011). Gao and Liang (2013) provide theoretical support for the argument that growth firms are endogenously opaquer than value firms. Holding all else constant, growth firms disclose less information than value firms and, therefore, have greater information asymmetry.

We separately split our sample into two equal subsamples based on firm age and book-to-market ratio. Firms that are younger (older) than the median in our sample are classified as young (old) firms, and those that have a higher (lower) book-to-market ratio than the median are classified as growth (value) firms. If the EDGAR implementation reduces retail investors' overconfidence by increasing the availability of firms' fundamental information, we expect the effect to be stronger among young and growth firms, which have high information asymmetry.

To test our predictions, we repeat our baseline analysis and the stacked diff-in-diff analysis using the subsamples. Table 1.5 shows the results for the young and old companies separately. Panel A presents the regression results with trading volume as the dependent variable. The first four columns (columns (1) to (4)) show the regression coefficients and corresponding t-statistics for young firms and the last four columns (columns (5) to (8)) for the old firms. We run both baseline staggered diff-in-diff regressions and stacked diff-in-diff regressions. We find that the effect of the EDGAR implementation on trading volume is more pronounced for young firms than for old firms. In baseline regressions, the coefficient on $Post$ is -0.114 in column (1) and -0.037 in column (5) with t- statistics of -4.05 and -1.43, respectively. The results indicate a significant decrease in trading volume by 10.4% of its standard deviation for young firms after the EDGAR implementation at the 1% level, whereas the effect is statistically nonsignificant for old firms (3.2%). The difference between these two coefficients is also statistically significant, as indicated by the p-value of 0 in the last two rows of Panel A. After including control variables, the coefficients on $Post$ are statistically significant for both young and old firms at the 1% level for young firms and the 10% level for old firms. Furthermore, with control variables in columns (2) and (6), trading volume decreases by around 11.4% of its standard deviation for young firms after the EDGAR implementation but only decreases by 4.1% for old firms. The difference between the two groups is statistically significant, indicating a larger trading volume decrease for young firms compared to old firms following the EDGAR implementation. We also find consistent results when using the stacked diff-in-diff regression as shown in columns (3), (4), (7) and (8). The coefficients on $Post$ are statistically significant for young firms at -0.125 and -0.116, whereas nonsignificant for old firms.

[Insert Table 1.5 here]

Panels B of Table 1.5 presents the regression results with dollar volume as the dependent variable. The results in general are consistent with the findings from the trading volume regressions in that the effects are more pronounced for young firms than old firms, especially for the staggered diff-in-diff regressions. For instance, the coefficient on $Post$ is -0.189 (t-stat = -4.67) in column (2) for young firms and -0.088 (t-stat = -1.85) in column (6) for old firms and the difference in coefficients is statistically significant at the 1% level. Although the results are weaker in the stacked diff-in-diff regression, the overall pattern suggests a greater decline in dollar volume for young firms than the old firms following the EDGAR implementation. For trading frequency (*Freq*), reported in Panel C the coefficients on $Post$ are significant and negative for both staggered and stacked diff-in-diff regressions for young firms as shown as shown in the columns (1) – (4).

In contrast, three coefficients on $Post$ are statistically nonsignificant for old firms as shown in columns (5), (7), and (8), and only significant at the 10% level for the staggered diff-in-diff regression with control variables as shown in column (6). On average, trading frequency decreases by around 10% of its standard deviation for young firms as shown in column (1) and 3.7% for old firms in column (5), which is nonsignificant (t-stat = -1.63).

Overall, the findings in Table 1.5 support our prediction that the impact of the EDGAR implementation on overconfidence measured by trading volume, dollar trading volume, and trading frequency is stronger for firms with higher information asymmetry measured by firm age.

We further test the effect of the EDGAR implementation on investors' overconfidence for growth versus value firms. The regression results of the staggered and stacked diff-in-diff approaches are reported in Table 1.6. Panel A presents the results with trading volume as the dependent variable. In columns (1), (2), (5), and (6), we present the staggered diff-in-diff regression results of Eq. (1.3), while columns (3), (4), (7), and (8) display the stacked diff-in-diff regression results of Eq. (1.4). For growth firms, the regression results indicate a decrease in trading volume by approximately 7.6% and 14.7% of its standard deviation in the staggered and stacked diff-in-diff regressions, respectively, as shown in columns (1) and (3). On the other hand, for value firms, the decrease in trading volume is smaller by around 7% and 5.7% of its standard deviation, as displayed in columns (5) and (7). Additionally, the differences between the coefficients for growth and value firms are statistically significant.

[Insert Table 1.6 here]

Panels B and C of Table 1.6 present the regression results when the dependent variables are dollar volume and trading frequency, respectively. In Panel B, when no control variables are included, the decrease in dollar volume is around 5.6% of its standard deviation for growth firms and 5.5% for value firms using the staggered diff-in-diff regression. However, the difference between the two coefficients is not statistically significant. When control variables are added, the decrease in dollar volume for growth firms is significantly larger than that for value firms, as shown in column (2). In the stacked diff-in-diff regression, we observe that the decrease in dollar volume is more pronounced for growth firms compared to value firms. The difference between the coefficients is statistically significant as indicated in columns (3) and (4). For example, the decrease in dollar volume is approximately 11.4% of its standard deviation for growth firms (*t*-stat = -3.08) in column (3), whereas it is 3.3% for value firms in Column (7) and not statistically significant. These findings suggest that the EDGAR implementation has a larger impact on reducing dollar volume for growth firms compared to value firms. We find consistent results for trading frequency in Panel C. The decrease in trading frequency is around 8.7% of its standard deviation (*t*-stat = -3.59) for growth firms and 6.2% (*t*-stat = -3.00) for value firms. The difference in coefficients between growth and value firms is statistically significant.

In conclusion, our findings support the hypothesis that the EDGAR implementation leads to a significant decrease in trading activities specifically for firms with high information asymmetry. This decrease can be attributed to the reduced costs incurred by investors when acquiring fundamental information about these firms.

1.4.3 Trading performance

In addition to analyzing trading activities, another way to measure investors' overconfidence is through their trading performance. Previous studies show that overconfident investors tend to overestimate their ability to process information, leading to investment mistakes (Odean, 1999). As a result, the stocks sold by overconfident investors are likely to systematically outperform the stocks purchased by them. Kumar (2009) measures the overconfidence of investors using the difference between stock returns following sells and purchases in each year. Following this measure, we calculate the post-trade sell-buy return differential (*PTSBD*) at the stock level in a quarter using Eq. (1.1) and Eq. (1.2). $PTSBD_{i,t}$ captures the difference between returns for stock i following sells and purchases of retail investors in quarter t. A larger (more positive) $PTSBD_{i,t}$ indicates that returns earned in 30 days following sells are higher than returns in 30 days following purchases. We expect that investors' overconfidence, measured by *PTSBD*, decreases on average after the EDGAR implementation.

We run both baseline staggered and stacked diff-in-diff regressions using Eq. (1.3) and Eq. (1.4) in which the dependent variable is *PTSBD*. Table 1.7 reports the regression coefficients and corresponding t-statistics. Columns (1) and (2) of Table 1.7 present the regression results of Eq. (1.3). In column (1), the coefficient on $Post$ is -31.13, with a t-statistic of -2.42. After including control variables that could affect investor overconfidence, the coefficient decreases to -26.14 as shown in column (2) but remains significant at the 5% level (t-stat = 1.97). These negative coefficients indicate a significant decrease in the difference between post-trade returns following sells and purchases after the implementation of EDGAR. The coefficient on $Post$ is also economically meaningful. For example, the coefficient of -31.13 indicates that, on average, the 30-day post-trade dollar returns following sells relative to those following purchases in the same quarter decrease by $31.13. Considering that the mean and standard deviation of *PTSBD* are -12.3 and 703.3, respectively, this decrease represents approximately 2.5 times of the mean and 4.4% of the standard deviation of *PTSBD* after the EDGAR implementation. We find consistent results in columns (3) and (4) using the stacked diff-in-diff regression. The corresponding coefficients on $Post$ are -69.88 and -73.24 for regression models with and without control variables, with t-statistics of -2.11 and -2.24, respectively. In conclusion, our analysis demonstrates that the 30-day post-trade dollar returns following sells for retail investors decrease compared to those following purchases after the implementation of the EDGAR system. This finding suggests a decrease in overconfidence among retail investors.

[Insert Table 1.7 here]

1.5 Robustness checks

In this section, we conduct further analyses to check the robustness of our main results regarding the impact of the EDGAR implementation on overconfidence of retail investors.

1.5.1 Excluding the first implementation phase

We exclude the firms that joined the EDGAR system in the first phase to address two concerns. The first concern relates to our assumption that the implementation of EDGAR was mandatory. Although the SEC planned to require all public companies to join the EDGAR platform, the first phase included companies that voluntarily chose to participate in the EDGAR system during its pilot period in the 1980s. Before the SEC began the mandatory EDGAR implementation for all U.S. public firms, it called for volunteers to disclose their information online. These voluntary companies were included in the first implementation phase, as recorded by SEC Release No. 33-6977. Therefore, the firms in phase one may not have been randomly assigned to that phase by the SEC, which conflicts with the randomized assignment assumption of the diff-in-diff analysis.

The second concern is related to the cost of acquiring financial statements using EDGAR. In the year-quarter when the first phase became effective (1993Q4), investors had to pay fees to access the financial statements available on EDGAR. Free access began on January 17, 1994, when Internet Multicasting Service, a nonprofit organization, and New York University made EDGAR filings available for free to Internet users (Liu, 2019; Gao and Huang, 2020). Therefore, the effect of the decrease in information acquisition costs may have been delayed because there was still a significant cost to access firm filings on EDGAR during the first implementation phase in the fourth quarter of 1993.

We address these two concerns by rerunning Eq. (1.3) and Eq. (1.4) with the firms in the first phase removed from the sample. The results are shown in Table 1.8. In Panel A, we report the regression results for the baseline staggered diff-in-diff analysis using Eq. (1.3). Panel B shows the results for the stacked diff-in-diff regression using Eq. (1.4). Columns (1) to (6) of Table 1.8 present the regression results when the dependent variables are trading volume, dollar volume, trading frequency and $PTSBD$. Across these columns, we consistently observe significantly negative coefficients for these trading-related measures, indicating a decrease in trading volume and frequency after the EDGAR implementation. In economic terms, all else being equal, trading volume decreases by approximately 8.6% of its standard deviation, dollar volume by around 6.4% of its standard deviation, and trading frequency by approximately 8.7% of its standard deviation (as shown in columns (2), (4), and (6) in Panel A, respectively). Columns (7) and (8) show the staggered regression results when the dependent variable is the $PTSBD$ without and with controls, respectively. The corresponding coefficients on $Post$ are -33.06 and -27.69 with t-statistics of -2.70 and -2.16. These coefficients indicate that, after the EDGAR implementation, the 30-day dollar returns following sells relative to purchases decrease by \$33.06 and \$27.69, respectively. The decrease in the former is approximately 2.69 times the mean and 4.7% of the standard deviation of $PTSBD$.

[Insert Table 1.8 here]

Panel B of Table 1.8 presents the consistent results using the stacked diff-in-diff regressions. In column (1), the results show that trading volume decreases by approximately 12.6% of its standard deviation following the EDGAR implementation. Similarly, in columns (3) and (5), dollar volume and trading frequency decrease by 8.5% and 9.9% of their standard deviations, respectively, with statistical significance at the 1% level. When examining the *PTSBD* as the dependent variable, the coefficient on $Post$ is -68.9 in column (7) without controls and -71.96 in column (8) with controls, with t-statistics of -2.05 and -2.14, respectively. These coefficients indicate a \$68.9 or \$68.9 decrease in the 30-day returns following sells relative to purchases after the EDGAR implementation. The results remain consistent with our main finding that the increase in information availability reduces the overconfidence of retail investors. Overall, the exclusion of firms from the first phase of the EDGAR implementation does not alter our main results, supporting the conclusion that increased information availability through the EDGAR system leads to reduced overconfidence among retail investors.

1.5.2 Pre-event parallel trend assumption and falsification test

To ensure the parallel trend assumption required for the diff-in-diff analysis, we follow the approach used by Gao and Huang (2020) and conduct tests using pseudo-event dates. The pseudo-events are set to two years before the actual effective date of the EDGAR implementation. Accordingly, the $Post$ indicator is defined according to the pseudo-event dates. If quarter t is after the pseudo-event date, then $Post$ equals one; otherwise, it equals zero. We discard the observations if quarter t is after the actual event dates, meaning that none of the firms in this test actually join the EDGAR system. We then rerun the baseline regression, Eq. (1.3), using the newly defined $Post$ indicator.

The results are shown in Panel A of Table 1.9. The coefficients on $Post$ are nonsignificant for all dependent variables, including trading volume, dollar volume, trading frequency, and *PTSBD*. These findings indicate that there is no substantial decrease in overconfidence before the

EDGAR implementation, supporting the parallel trend assumption required for the diff-in-diff analysis.

[Insert Table 1.9 here]

In addition to using the pseudo-events preceding the actual EDGAR implementation to test the pre-event parallel trend assumption, we also use the pseudo-events occurring two years after the actual implementation dates to conduct the falsification test, following Gao and Huang (2020). If the decrease in overconfidence is related to the decrease in information acquisition costs, we would expect a nonsignificant decrease in overconfidence around the pseudo-events. The $Post$ indicator equals one if quarter t is after the pseudo-event date; otherwise, it equals zero. We discard the observations if quarter t is before the actual event dates, indicating that all firms in this test have already joined the EDGAR system. The results, presented in Panel B of Table 1.9, show no significant decrease in trading volume and dollar value during the pseudo-event periods. We observe a marginal increase in trading frequency after the artificial event times, significant at the 10% level without control variables in columns (5) and at the 5% level with controls in column (6). For the $PTSBD$, we find nonsignificant increases in the coefficients on $Post$, supporting our main results that the decrease in overconfidence can be explained by the decrease in information acquisition costs resulting from the EDGAR implementation.

1.5.3 Diff-in-diff analysis by each implementation phase

To address concerns about the potential influence of specific phases of the EDGAR implementation on the results of the staggered diff-in-diff analysis, we conducted a baseline analysis for each phase. The results are presented in Table 1.10. In columns (1) and (2), we show the regression results with trading volume as the dependent variable, both without and with control variables. For firms in all first six phases, we find negative coefficients for the $Post$ variable.

However, the magnitude of the decrease in trading volume varies across the phases. Four out of the six coefficients are statistically significant. In phases 2, 3, 4, and 5, the coefficients on $Post$ are -0.22, -0.20, -0.21, and -0.10, respectively, with corresponding t-statistics of -2.54, -3.93, -3.94, and -2.52. Columns (3) and (4) display the regression results when the dependent variable is dollar volume and it is trading frequency in column (5) and (6). The negative coefficients indicate a decrease in dollar volume and trading frequency for firms in all six phases, supporting our main findings. For the $PTSBD$ analysis, the regression results in each phase yield negative coefficients. This suggests that overconfidence decreases after the EDGAR implementation. Notably, the treatment effect appears stronger in phases 2 and 5, as indicated by the magnitudes of the coefficients and their statistical significance levels. Taken together, these results indicate that the observed decreases in retail investors' overconfidence after the EDGAR implementation are not driven by a specific implementation phase. The findings remain consistent across different phases, supporting the robustness of our main conclusions.

[Insert Table 1.10 here]

1.6 Conclusion

We examine how information acquisition costs affect retail investors' overconfidence using the EDGAR implementation as an exogenous shock to information acquisition costs. Following the EDGAR implementation, which mandated firms to transition from paper-based financial disclosures to digital formats, retail investors could acquire firms' financial statements online without a fee. Consequently, the decrease in information acquisition costs provided retail investors with greater access to fundamental information. With more financial data available, it becomes more difficult for retail investors to construct a highly coherent but false story for making investment decisions. To measure the overconfidence of retail investors, we analyze transaction data from a large U.S. brokerage during the period from 1991 to 1996. Using staggered diff-in-diff analysis, we find that the overconfidence decreases substantially after the EDGAR implementation. Our results are confirmed by a stacked diff-in-diff analysis, which avoids the potential estimation bias caused by the heterogeneity between groups in the traditional staggered diff-in-diff analysis. Our results remain unchanged after a battery of robustness checks.

To better understand the mechanisms underlying retail investors' overconfidence, we divide the sample according to firms' information asymmetry. As retail investors are at a disadvantage in accessing information about firms, especially those of firms with high information asymmetry, we expect the EDGAR implementation to have a stronger effect on investors' overconfidence in the subsample of firms with high information asymmetry. We indeed find that the reduction in overconfidence among investors after the EDGAR implementation is stronger for young firms and growth firms, both of which exhibit high information asymmetry. In conclusion, this paper provides direct evidence that the degree of overconfidence changes with the information set held by investors, which adds to the understanding of the determinants of the overconfidence of retail investors.

Chapter 2 Earnings Guidance and Ex-ante Stock Price Crash Risk

2.1 Introduction

A large amount of accounting literature examines how stock price crash risk is determined by earnings information released by firms. Theoretical models point out a predictive relation between firms' information environment and stock price crash risk (Jin and Myers, 2006; Bleck and Liu, 2007). These models suggest that the likelihood of stock price crash is higher when the information transparency regarding firm-specific negative news is limited[9]. Empirically, using the realized skewness as the measure of stock price crash risk, studies find that there is a decrease in stock price crash risk after the improvement of firms' information environment (Hutton, Marcus and Tehranian, 2009)[10].

Management guidance, as one of the earnings information disclosure channels, provides around 55% of accounting-based information in explaining quarterly stock return variance (Beyer, Cohen, Lys and Walther, 2010). Previous studies examine both managers' incentives to provide earnings forecasts and the market response to the forecasts[11]. Through management guidance, managers release their forecasts of firms' future earnings to the public, reducing the information asymmetry between insiders and outsiders and increasing firms' information transparency (Coller and Yohn, 1997; Chen, Matsumoto and Rajgopal, 2011).

We examine whether management guidance affects investors' subjective stock price crash

[9] When opaque firms conceal negative earnings information, and insiders have limited capacity to absorb such information, it leads to numerous negative outliers in the left tail of stock return distributions, which are identified as stock price crash.
[10] Kim, Wang, and Zhang (2019) state that less readable 10-K reports are associated with high stock price crash risk. DeFond, Hung, Li and Li (2015) show that IFRS adoption decreases stock crash risk among nonfinancial firms if firms have a poor information environment. Such a decrease in stock crash risk could be attributed to credible changes in local GAAP after the IFRS adoption. Hsu, Wang and Whipple (2022) document that non-GAAP disclosure would increase future realized stock crash risk. They explain this positive relation as the non-GAAP reporting may convey optimistically biased information about firms' future earnings.
[11] See representative studies including Patell (1976), Penman (1980), Baginski, Conrad and Hassell (1993), Skinner (1994), Miller (2002), Hutton, Miller and Skinner (2003), and Rogers, Skinner and Van Buskirk (2009).

risk in a short window. When managers disclose bad news, the probability of bad news accumulation decreases, leading to a reduction of future stock price crash risk. Therefore, we expect that investors' subjective stock price crash risk will decrease once investors notice such voluntarily disclosed bad news by managers. Releasing good news may also help decrease the information asymmetry, but there are concerns regarding the truthfulness of the good news that has been voluntarily disclosed through guidance. For instance, CEOs may make opportunistic voluntary disclosure decisions to maximize their stock option compensation (Noe, 1999), trading profits (Cheng and Lo, 2006) or to influence the sentiment-induced biases in investors' expectation (Bergman and Roychowdhury, 2008). In addition, Cotter, Tuna and Wysocki (2006) suggest that managers use the earnings guidance to manipulate analyst forecasts. Given the evidence that highlights investors' concerns about the credibility of guidance for good news, we expect that there is no significant decrease, or even an increase, in ex-ante stock price crash risk when managers release their positive forecasts.

We use option data to measure investors' perceptions of future stock price crash risk because options, particularly out-of-the-money (OTM) options, provide a forward-looking measure of market sentiment and expectations. The prices of these options reflect the collective beliefs and risk assessments of market participants regarding future movements in the underlying stock price. To capture investors' concerns about potential crashes in stock price, we use option-implied skewness as the proxy for investors' expectations of downside risk, as it measures the asymmetry of the implied distribution of future stock prices. When investors perceive a higher risk of significant downward movements in the stock price, they typically demand higher premiums for put options (which provide protection against declines) relative to call options. This demand imbalance results in a skewed distribution, indicating heightened concern about downside risk.

Using option data from 1996 to 2022, we compute ex-ante skewness and variance from OTM options with various maturities and interpolate the implied skewness and variance to 30-, 60-, and 91-day maturities, following the methodology of Schneider, Wagner and Zechner (2020). The skewness is defined as the sum of upper and lower skewness, corresponding to the positive and negative parts of the distribution, capturing the risk-neutral skewness of returns. This ex-ante skewness enables us to measure investors' expectations of future stock price crash risk over these time horizons and investigate how daily ex-ante skewness changes in the short period surrounding the release date of management forecasts.

We first validate the implied ex-ante skewness measure. The results show that option-implied skewness contains information about future realized stock price crash risk. The implied skewness using options of available maturities interpolated to maturities of 30, 60, and 91 days can predict the realized skewness in the future 30, 60, and 91 days, where the realized skewness is calculated using daily stock returns, indicating the ex-ante skewness calculated from option data contains information of future realized stock price crash risk.

Next, we examine the change in implied skewness as the reversed measure of investors' subjective beliefs about crash risk around the disclosure date of management guidance. To ensure the impact of earnings announcements does not confound our results, we specifically exclude management guidance announcements made in conjunction with earnings announcements.[12] Our findings reveal that, on average, the ex-ante skewness during the two days following the report date of quarterly management guidance is significantly higher (i.e., less negative) than the ex-ante

[12] We exclude the earnings announcements because previous studies show that implied volatility tends to increase in the period leading up to the earnings announcement date and decrease thereafter as earnings announcements follow predetermined schedule (Patell and Wolfson, 1979; Patell and Wolfson, 1981; Isakov and Pérignon, 2001; Rogers, Skinner and Van Buskirk, 2009). Such pattern of implied volatility change may affect the implied skewness change around the earnings announcements. Given that we use option-implied skewness as a measure of investors' belief in future stock price crash risk, the increase in implied skewness (or less negative skewness) could potentially be attributed to the decrease in implied volatility following earnings announcements.

skewness during the two days preceding the guidance disclosure. Since we use option-implied skewness as a reversed measure of expected stock price crash risk, the increase in ex-ante skewness after the disclosure management guidance indicates a decrease in investors' expectations of stock price crash risk.

Theoretical models suggest that the decrease in stock price crash risk after the firms' information disclosure could be attributed to the disclosure of bad news. Models proposed by Jin and Myers (2006) imply that managers have the incentive to hide or conceal bad news due to career concerns. When accumulated bad news reaches a critical point and is released all at once, it can trigger extreme stock price declines. Based on the model implication, we should expect to observe a decrease in investors' belief of future stock price crash risk concentrated around the management guidance that releases bad news.

To examine this prediction, we categorize management guidance disclosures based on whether firms report bad or good news. We use the "guidance_code" variable provided by the IBES management guidance database, which captures forecast earnings relative to analyst consensus. According to the variable's definition, we define bad (good) news where firms' guidance is below (above) the mean of analysts' estimates. Consistent with the theory, we find that investors' subjective stock price crash risk decreases after firms announce bad news. However, when firms disclose good news, investors' expected stock price crash risk increases, with a smaller magnitude. This observation aligns with previous studies suggesting that bad news forecasts are generally more informative than good news forecasts, with a larger market response (Hutton, Miller and Skinner, 2003)[13]. For the placebo event, we do not observe a significant change in ex-

[13] It is possible that option traders are already aware of the news that may lead to poor (good) performance of firms before the earnings guidance is publicly released. If this is the case, the increase (decrease) in ex-ante skewness observed after the bad (good) news disclosure might be understated, as we conduct our analysis within a short window

ante skewness when analysts' consensus falls within the expected forecast range reported by the managers (neutral news).

The decrease in subjective stock price crash risk following bad news is related to the increase in information transparency with a decreased probability of hiding the bad news. With a higher level of informativeness in the management guidance disclosure, we expect a more pronounced decrease in stock price crash risk. To proxy the level of informativeness, we use the periodicity of the management forecasts and an indicator showing whether multiple guidance measures are reported[14]. Earnings forecasts for the next fiscal quarter are likely to provide more accurate predictions than those for the next fiscal year due to the shorter forecasting horizon and the availability of more current information. Additionally, having predictions for multiple accounting variables in each management forecast press should provide more informative insights compared to forecasts that contain predictions for only one variable. Using regression models with interaction terms between bad news and level of informativeness, we find that the decrease in ex-ante crash risk following bad news is more significant when quarterly forecasts or multiple guidance measures are reported. However, we did not observe statistically significant impacts of the level of informativeness on ex-ante crash risk following good news or neutral news. This implies that there may be alternative explanations for the increase in ex-ante crash risk following good news.

Next, we investigate the underlying mechanisms that explain the opposite changes in ex-ante skewness observed following the disclosure of bad news versus good news. The decrease in subjective stock price crash risk after the bad news disclosure can be attributed to the decrease in information asymmetry (Bleck and Liu, 2007). With more bad news disclosed by managers and

---

(2 days before and 2 days after the earnings guidance). The real changes in investors' expectations of downside (upside) risk might be more pronounced than what our results suggest.

[14] The guidance measure includes sales, return on equity, return on assets, pretax income, net income, gross margin, earnings per share, EBITDA, dividends per shares and capital expenditure.

less hidden, the probability of accumulating undisclosed bad news decreases, leading to less extreme stock price drops in the future when the hidden bad news is eventually uncovered. If the decrease in subjective stock price crash risk is attributed to the decrease in information asymmetry, we should expect to observe a larger decrease for firms with high information asymmetry. By using the number of analyst coverage as a proxy for information asymmetry, we find that investors' perceived crash risk decreases more among firms with lower analyst coverage. The result indicates that the decrease in information asymmetry can potentially explain the decrease in ex-ante stock price crash risk after the bad news disclosure of management guidance.

For good news disclosure, the increase in investors' belief of future stock price crash could be explained by the agency problem. The agency model proposed by Hermalin and Weisbach (2012) implies that managers might disclose biased information based on their career concerns, distorting investors' perception of firm value. Earnings guidance, as a voluntary disclosure, can be used opportunistically by managers to generate biased expectation for their own benefits (Rogers and Stocken, 2005; Cheng and Lo, 2006). Releasing good news may be a strategy used by managers to hide bad news to gain short-term benefits. However, stockholders may sue when there is a large decline in stock price on earnings announcement day if managers fail to promptly disclose bad news promptly. Previous studies find that one of the reasons for managers to voluntarily disclose bad news is the fear of litigation risk (Skinner, 1994; Skinner, 1997). Cheng and Lo (2006) also find that managers selectively exploit voluntary disclosure opportunities to maximize their trading profit when litigation risk is sufficiently low. We investigate whether the changes in ex-ante skewness vary systematically with the firms' litigation risk. Investors may expect a higher future stock price crash risk when they suspect that the good news announced by managers is a tool to hide bad news. When litigation risk is higher, managers are less likely to disclose good news opportunistically, making the good news more credible. We find that the increase in expected stock price crash risk following good earnings news guidance becomes smaller for firms with higher litigation risk, indicating that the increase in expected stock price crash risk following good earnings news guidance could be attributed to weak credibility of good news disclosures.

Furthermore, we investigate whether the change in investors' belief in future stock price crash risk around the disclosure of management guidance can predict future realized stock price crash. We find that the decrease in ex-ante stock price crash risk, which reflects expectations for the next 30 days, can predict realized stock price crashes in the next month, controlling for the past levels of ex-post and ex-ante stock price crash risk. However, we do not observe statistically significant predictive power from the change in expectations for the next 60 and 91 days.

Our paper contributes to the studies related to earnings information opacity and stock price crash risk. The majority of previous studies use the realized stock price crash following Chen, Hong and Stein (2001), which is an ex-post measure and commonly requires one-year weekly stock returns for estimation. Consequently, previous studies focus on how changes in the stock price crash in the current year could be affected by the changes in firms' information environment in previous year. Empirical evidence suggests that if firms have a higher tendency to hide negative earnings news and have low information transparency, stock price crash risk will be larger in the future (Kim, Wang and Zhang, 2019; Li and Zhan, 2019). However, little empirical evidence shows how ex-ante skewness changes in a short window around firms' disclosure of earnings information.[15] Our paper contributes to this topic by providing empirical evidence using ex-ante

[15] A few studies use the ex-ante annual skewness as the measure of stock price crash risk. Kim and Zhang (2014) and Kim, Li, Lu and Yu (2016) use the option-implied volatility smirk, which is calculated as the difference between the implied volatility of an out-of-the-money (OTM) put option and that of an at-the-money (ATM) call option on the same day. However, they do not examine the change in expected skewness within a short window around earnings disclosure dates. Instead, both papers use the mean of the daily implied volatility smirk over 12 months ending three months after the fiscal year-end in the empirical setting as the measure for expected crash risk.

skewness, which captures investors' belief of future stock price crash risk. Using the more frequently updated daily skewness data, we find investors tend to lower their expected short-term stock price crash risk after managers issue bad earnings guidance. This approach allows us to capture more immediate changes in investor sentiment and expectations, enhancing our understanding of the dynamics between earnings disclosures and stock price crash risk.

Our paper also contributes to the literature on the impact of management earnings forecasts. Previous literature points out that management earnings forecasts help to increase the frequency of information disclosure beyond scheduled earnings announcements. Moreover, managers are found to have preferences for disclosing negative news related to future earnings. Rogers, Skinner and Van Buskirk (2009) highlight that management earnings forecasts increase short-term implied volatility when managers disclose bad news sporadically. Theoretical models from behavioral perspective imply that managers strategically disclose bad news to cater to loss-averse investors and guide down investors' expectations of future earnings (Huang, Piotroski and Xie, 2023). Our study contributes to this literature by documenting that investors' expected stock price crash risk decreases after managers release bad news and increases following good news. With the negative news disclosed by managers, the likelihood that firms are hiding bad news and will release it all at once in the future decreases. This finding provides empirical support for the idea that management earnings forecasts play a crucial role in shaping investors' expectations about future stock price crash risk, particularly by enhancing transparency and reducing information asymmetry when bad news is disclosed.

2.2 Literature review

2.2.1 Estimating stock price crash risk

Stock price crash refers to the extremely negative stock returns that can be captured by the higher moment of the stock return distribution. The literature argues that such outliers in the left-tail of return distributions could be attributed to the accumulation of bad news related to future earnings. If managers hide the negative earnings information on purpose, there will be an asymmetric stock return response to bad news and good news. The magnitude of positive returns corresponding to good news is smaller than that of negative returns to bad news (Kothari, Shu and Wysocki, 2009). When the accumulated bad news becomes publicly available at once, there would be a large drop in stock price known as a stock price crash.

Previous literature uses firm-specific ex-post skewness as the measure of stock price crash risk, which is first proposed by Chen, Hong and Stein (2001). Realized skewness is estimated every year at the firm level as the third moment of return distribution using the weekly stock returns adjusted by the lead and lag market returns. Negative skewness indicates the stock return distribution is skewed to the left. To capture the realized skewness, the stock crash dummy variable is used as an alternative measure. The crash dummy equals one if the stock return falls at least 3 standard deviations below its mean value in a given year (Hutton, Marcus and Tehranian, 2009). In addition to estimating the realized skewness, Kim and Zhang (2014) use the implied volatility smirk as the difference between the implied volatility of OTM puts and that of ATM calls as the measure for implied skewness. Such implied skewness is estimated during a 12-month period three months before the fiscal year ends, serving as an annual proxy for each stock.

2.2.2 Information disclosure and stock price crash risk

Previous studies examine the decisive factors of stock price crash risk from the perspective of financial disclosure. Jin and Myers (2006) establish theoretical models that explain the frequency of stock price crashes using the stock information environment. Their model emphasizes the importance of whether the importance of the transparency of a firm's information in explaining the occurrence of outliers in the left tail of future stock return distributions. With limited ability for insiders to absorb the negative bad earnings information, when the firms are more opaque, with greater information hidden, there is a higher probability for stock price crashes when the bad news becomes publicly available all at once in the future.

Based on this theoretical model, Hutton, Marcus and Tehranian (2009) first provide empirical evidence showing that opaque firms are more likely to experience stock price crashes, using accumulated accruals as a measure of earnings management. In addition to aggregate earnings management, DeFond, Hung, Li and Li (2015) show that changes in accounting reporting standards can also affect stock price crash risk. They find that firms whose information transparency improved after the IFRS adoption experience a significant decrease in future stock price crash risk. This effect is more pronounced in firms with poorer information environments and in countries where IFRS adoption is implemented more credibly.

Moreover, studies also focus on the information contained in the financial statements. Ertugrul, Lei, Qiu and Wan (2017) find that firms with larger sizes of 10-K files and more uncertain words in 10-Ks have greater future stock price crash risk. Using a similar setting but a different proxy, a modified version of the Fog Index, Kim, Wang and Zhang (2019) find that firms with less readable 10-K are associated with more negatively extreme returns. Additionally, Kim, Li, Lu and Yu (2016) highlight that financial statement comparability is associated with ex-ante stock price crash risk. By using the implied volatility smirk, defined as the difference between the implied volatility of an OTM put on a day and that of an ATM all on the same day, they find that the increased financial statement comparability decreases the steepness of the implied volatility, indicating a decreased stock price crash risk.

2.3 Sample and measurements

In this section, we discuss our sample selection, data source and the measures of ex-ante skewness.

2.3.1 Sample selection

Our sample includes the common equity share class of a U.S. corporation with share classes of 10 and 11 and stocks listed on the New York Exchange, American Stock Exchange, and Nasdaq Stock Market-National Market system with the exchange codes equal 1, 2, and 3. We exclude the financial services and utility firms with the Standard Industrial Classification (SIC) code between 6000 and 6900 and between 4900 and 4949 respectively, following the industry definition for 48 industry portfolios from Kenneth French's website. The sample period is from January 1996 to December 2022.

To estimate the ex-ante skewness and implied volatility of stocks, we obtain equity option data from OptionMetrics. We obtain financial statement data from Compustat, stock price data from the Center for Research in Security Prices (CRSP), and analyst coverage data from the Institutional Brokers' Estimate System (I/B/E/S). Institutional ownership from 13F fillings by institutional investors. In order to examine the change in ex-ante skewness around the management guidance, we obtain the date on which the management earnings forecast is released from the IBES Management Guidance database (Rogers and Van Buskirk, 2013; Lu and Skinner, 2020). We include both quarterly and annual management guidance. To rule out the impact of earnings announcements in our analysis, we exclude the management guidance announced in conjunction with earnings announcements.

2.3.2 Ex-ante skewness

We estimate the ex-ante skewness as the standardized forward-looking proxy for investors' perceived stock price crash risk using the OTM stock option data following Schneider, Wagner and Zechner (2020). The skewness is calculated as the sum of upper and lower skewness, which captures the positive and negative part of return distribution respectively, as shown in Eq. (2.1)

$$SKEW_{t,T} = \frac{6}{p_{t,T}}\left(\int_{F_{t,T}}^{\infty} \log\left(\frac{K}{F_{t,T}}\right)\frac{\sqrt{\frac{K}{F_{t,T}}}C_{t,T}(K)}{K^2}\mathrm{dK} - \int_{0}^{F_{t,T}} \log\left(\frac{F_{t,T}}{K}\right)\frac{\sqrt{\frac{K}{F_{t,T}}}P_{t,T}(K)}{K^2}\mathrm{dK}\right), \quad (2.1)$$

$$VAR_{t,T} = \frac{2}{p_{t,T}}\left(\int_{0}^{F_{t,T}} \frac{\sqrt{\frac{K}{F_{t,T}}}P_{t,T}(K)}{K^2}\mathrm{dK} + \int_{F_{t,T}}^{\infty} \frac{\sqrt{\frac{K}{F_{t,T}}}C_{t,T}(K)}{K^2}\mathrm{dK}\right), \quad (2.2)$$

where $F_{t,T}$ is the forward price of the stock with the contracted date as time t and delivery at time T. $C_{t,T}(K)$ and $P_{t,T}(K)$ capture the prices of European call and put options with the strike price K respectively. $p_{t,T}$ is the price of zero-coupon bond with maturity T. The skewness is measured at the stock-day level. We further estimate ex-ante variance shown in Eq. (2.3). We calculate the ex-ante skewness and ex-ante variance with options of available maturities and interpolate skewness and variance to maturities of 30, 60 and 91 days, with N = 30, 60, and 91. The ex-ante skewness used in empirical analysis is scaled by ex-ante variance, shown in Eq. (2.3) below. The sign of $stdSKEW_{i,t,T}$ depends on the relative prices of OTM put and OTM call options. The increase in negative $stdSKEW_{i,t,T}$ (less negative) captures the decrease in ex-ante crash risk.

$$stdSKEW_{i,t,T} = \frac{SKEW_{i,t,T}}{VAR_{i,t,T}^{\frac{3}{2}}}. \quad (2.3)$$

To examine the change in skewness around the report date t of earnings information, we calculate the change in average skewness from two days before the information report date t to two days after date t, shown in Eq. (2.4)

$$DSKEW_{i,t,T} = \frac{stdSKEW_{i,t+1,T} + stdSKEW_{i,t+2,T}}{2} - \frac{stdSKEW_{i,t-1,T} + stdSKEW_{i,t-2,T}}{2}, \quad (2.4)$$

where $DSKEW_{i,t,T}$ captures the change in skewness around the guidance information disclosed at date t for stock i using the options with maturity T.[16] $DSKEW_{i,t}$ equals zero means no skewness change. A positive $DSKEW_{i,t,T}$ means skewness increases after the disclosure day t, indicating a decrease in expected stock price crash risk.

2.3.3 Control variables

Following the previous literature, we introduce the control variables that may affect the stock price crash risk. We control systematic risk and business risk using the market beta (*Beta*) and financial leverage (*Leverage*). Following Fama and French (1992), we estimate the market beta for each stock using its monthly returns over the prior 60 months. Leverage is the ratio of the total liabilities over the total assets. Firm performance is measured by using *ROA*, as stated by Hutton, Marcus and Tehranian (2009) that firms' operating performance is negatively related to the stock price crash risk. Book-to-market ratio (*B/M*), market capitalization (*Size*), stock turnover (*Turnover*) are also controlled, as growth firms, large firms and stocks with higher turnover are more likely to have stock price crashes (Chen, Hong and Stein, 2001; Kim and Zhang, 2016). We also control the contemporaneous daily stock returns (*Ret*) due to price level variations induced by earnings forecasts[17].

Firm size (*Size*) is computed as the natural logarithm of the market value of equity, calculated in million dollars. *B/M* is the firm's book equity at the end of the previous fiscal year, divided by

---

[16] When calculating the two averages in the equation, if $skew_{i,t+1}$ is missing then $\frac{skew_{i,t+1}+skew_{i,t+2}}{2}$ returns $skew_{i,t+2}$. If both $skew_{i,t+1}$ and $skew_{i,t+2}$ is missing, then $\frac{skew_{i,t+1}+skew_{i,t+2}}{2}$ returns missing.

[17] See Appendix A.

the market value of equity at the end of December of the previous year. *Turnover* is measured as the daily average number of shares traded in a quarter divided by the number of shares outstanding, shown as a percentage. We adjust the number of shares for Nasdaq stocks to address the double counting issue following Gao and Ritter (2010). Following Ang, Hodrick, Xing and Zhang (2006), monthly stock idiosyncratic volatility (*IVOL)* is computed as the standard deviation of the daily residuals in a month from the Fama–French (1993) three-factor model.

We also include analyst coverage and percentage of institutional shareholders to control the impact of cross-sectional variation in information asymmetry. Analyst coverage (*ACov*) is calculated as the natural logarithm of one plus the average number of analysts covering a stock in each quarter, and institutional holdings (*INST*) is measured as the natural logarithm of one plus a stock's institutional ownership scaled by shares outstanding in each quarter. To examine the impact of good and bad guidance disclosures separately, we define good news (*Good*) as a dummy variable that equals one if a firm's guidance is above the mean of analysts' estimation and zero otherwise. Bad news (*Bad*) is defined similarly. We also define neutral news (*Neutral*) when the mean of analysts' estimates falls within the expected range reported by the managers with no surprises. We consider these instances as placebo events, serving as a reference point for comparison. Disclosures are classified according to "guidance_code" variable in IBES management guidance database, which captures the guidance relative to consensus[18]. *Multi_type* is defined as a dummy variable that equals one when more than one type of guidance measures is reported, and zero otherwise.

[18] According to the definition of "guidance_code" in IBES, when both *Good* and *Bad* dummies are zero, it means either the company has provided guidance but not specified whether they will meet, beat or miss the street or the company is expected to meet earnings for the period indicated. For example, if the earnings forecast states that the expected EPS in next quarter is more than, slightly more than or significant more than 5 and the available mean of analyst estimation of EPS is 6, then both *Good* and *Bad* are zero. However, if the mean of analyst estimation is 4, then *Good* is one and *bad* is zero.

The summary statistics of the control variables and variables of interest are presented in Table 2.1. Panel A shows the change in ex-ante skewness and the classification variables for subsamples for quarterly forecasts, whereas Panel B for annual forecasts. In Panel A for quarterly forecasts, on average, the ex-ante skewness increases (becomes less negative) two days after the disclosure day from two days prior, as shown by the mean of $DSKEW$. The means of $DSKEW$ are 0.04, 0.03, and 0.03 when interpolating ex-ante skewness to maturities of 30, 60 and 91 days. In this subsample, 19% disclosures are bad news and 13% are good news, and 25% are neutral news. The remaining news (43%) are the ones where the analysts' consensus and the managers forecasts are not comparable[19]. In Panel B for annual forecasts, on average, the ex-ante skewness decreases (becomes more negative). The means of $DSKEW$ are -0.11, -0.04 and -0.02 when interpolating ex-ante skewness to maturities of 30, 60 and 91 days. In this subsample, there are slightly more good news disclosures (15%) than bad news disclosures (13%). In the full sample reported in Panel C, the ex-ante skewness decreases on average, indicating an increase in ex-ante crash risk. The means of $DSKEW$ are -0.06, -0.02, and -0.01 when interpolating ex-ante skewness to maturities of 30, 60 and 91 days.

[Insert Table 2.1 here]

Panel D of Table 2.1 reports the summary statistics for the control variables in the full sample. The mean of stock returns on the day of disclosure is -0.22%, indicating a decrease in stock price after the earnings forecast disclosure. The mean and standard deviation of ROA are 0.01 and 0.04. The average financial leverage and B/M are 0.21 and 0.52, respectively. The natural logarithm of the market value of the stocks in million dollars (*Size*) is 8.00. The mean of turnover is 1.16,

[19] For example, the analyst estimations are not available or there is no numeric forecast reported in the management guidance.

indicating that on average, 1.16% of the outstanding shares of the firms change hands in a day. The average market beta is 1.33. The mean and median of idiosyncratic volatility are 1.83 and 1.52 during a month. The mean of institutional ownership is 0.51, indicating that the average percentage of shares held by institutional investors is 67%. The mean of analyst coverage is 2.31, suggesting that the average number of analysts following the stocks is 10.07. Around 33% of reports are quarterly management guidance discloses and 18% of the disclosures report more than one type of guidance measure.

2.4 Empirical results

In this section, we examine whether and how investors' belief about future stock price crash risk changes around the management guidance disclosure and whether the change in investors' subjective crash risk can predict future stock price crashes.

2.4.1 Validity check of the ex-ante skewness measure

Before conducting the empirical analysis using the option-implied ex-ante skewness as the proxy for investors' subjective stock price crash risk, we first examine the validity of the ex-ante skewness measure. If option-implied skewness is estimated following Eq. (2.3), it should contain information about the ex-post skewness and therefore predict the future realized stock price crashes. We run the following regression model at the daily level in our sample. If implied skewness is a valid measure for the future realized stock price crashes, β_1 is predicted to be positive.

$$rSKEW_{i,t+1,t+1+N} = \alpha + \beta_1 \times stdSKEW_{i,t,t+N} + \boldsymbol{\beta} \times \boldsymbol{X} + FE_i + FE_t. \qquad (2.5)$$

The dependent variable $rSKEW_{i,t,t+N}$ is the standard skewness estimated using the daily stock return of stock i from day t to day $t + N$. The main independent variable is $stdSKEW_{i,t,t+N}$, which is estimated following Eq. (2.3) using the options of stock i on day t interpolated maturities to N days. $\boldsymbol{X}$ represents the vector of the control variable. We also include the firm and day fixed effects for unobserved factors across firm and time. The standard errors are clustered by stock and day to account for time-series and cross-sectional correlation in the residuals.

The regression results are reported in Table 2.2. We estimate the ex-ate skewness using options with available maturities and interpolate skewness to maturities of 30, 60 and 91 days, shown in the first two, middle two and last two columns from left to right. Columns (1) and (2) show the results of the regression model in which the independent variable is ex-ante skewness estimated using options with available maturities and interpolated to maturities of 30, and the dependent variable is the skewness of next 30 days realized daily stock returns. The results are consistent with our prediction. The coefficients on ex-ante skewness are 0.017 and 0.014, respectively, without and with the control variables added and both are statistically significant at 1% level. The coefficient in column (1) indicates that a one-standard-deviation increase in ex-ante skewness predicts a 0.04 (=0.017×1.82) increase in ex-post skewness. We find consistent results using ex-ante skewness with maturity of 60 days and 91 days, as shown in columns (3) to (6). All coefficients on ex-ante skewness are significantly positive at the 1% level, indicating the implied skewness estimated following Eq. (2.3) can be used as a valid measure for investors' perceptions of future stock price crashes.

[Insert Table 2.2 here]

2.4.2 Ex-ante skewness trends within [-10,10] window of the event day

As implied by the theoretical models, the disclosure of bad news could decrease future stock price crash risk. We examine whether investors' beliefs about future stock price crashes are affected within a short window around the management guidance disclosure. In this section, we analyze the trend of option-implied skewness within a [-10, 10] window around the disclosure date.

We first plot the average ex-ante skewness, $stdSKEW_{i,t,t+N}$, estimated following Eq. (2.3) for bad news and good news separately in Figure 2.1. The management guidance is disclosed on day 0. With bad news disclosed by the managers, fewer bad news tends to accumulate and be revealed at once in the future, leading to a decrease in future stock crash risk. Plots 1A, 1B, and 1C present the trend of ex-ante skewness with maturities of 30, 60, and 91 days, capturing investors' expectations for the future 30, 60, and 91 days, respectively. From all three plots, we observe a sharp increase in ex-ante skewness on day 0 and day 1, the day after the guidance disclosure. Taking Figure 1A as an example, ex-ante skewness estimated using options with an interpolated maturity of 30 days jumps from -1.12 to -0.83 on the disclosure day and continues to increase to around 0.29 one day after the disclosure. Since ex-ante skewness is measured as the reverse measure of stock price crash risk, the increase in implied skewness reflects a decrease in subjective stock price crash risk. The trends of ex-ante skewness are more stable and parallel before and after the disclosure day when using ex-ante skewness with longer maturities, 60 and 91 days, as shown in plots 1B and 1C, reflecting longer period forecasts.

[Insert Figure 2.1 here]

In Figure 2.2, we illustrate the ex-ante skewness trends around the management guidance when good news is disclosed, defined as management guidance being above the mean of analysts' estimates. Similar to Figure 2.1, we present the ex-ante skewness trend with maturities of 30, 60, and 91 days, as shown in Plots 2A, 2B, and 2C. In all three plots, we observe a decrease in ex-ante skewness on the disclosure day. Take Plot 2A as an example. The option-implied skewness for the future 30 days decreases from around -1.69 to -1.90 on the disclosure day. The more negative ex-ante skewness indicates an increase in investors' belief in future stock price crashes once the good news is forecasted. In conclusion, Figures 2.1 and 2.2 reflect consistent results, where investors' belief in future stock price crash risk decreases (increases) with the guidance of more (less) bad news by managers.

[Insert Figure 2.2 here]

2.4.3 Ex-ante skewness change within [-2,2] management guidance

As presented in Figures 2.1 and Figure 2.2, the change in ex-ante skewness occurred immediately after the guidance was disclosed. We now focus on the change in skewness within the narrower [-2, 2] window in the following empirical analyses. We calculate the DSKEW following Eq. (2.4), which captures the change in ex-ante skewness before and after the guidance disclosure day. According to Eq. (2.4), a positive DSKEW implies an increase in ex-ante skewness after the disclosure day, representing a decrease in expected stock price crash risk in the corresponding time period. To examine this change in ex-ante skewness, we employ the following regression model.

$$DSKEW_{i,t,t+N} = \alpha + \beta_1 \times Bad_{i,t} + \beta_2 \times Good_{i,t} + \beta_3 \times Neutral_{i,t} + \boldsymbol{\beta} \times \boldsymbol{X} + FE_i + FE_{qtr} \quad (2.6)$$

Bad, $Good$ and $Neutral$ are dummy variables that equals one if the firm discloses bad news, good news and neutral news on the disclosure day, and zero otherwise. We predict that β_1 is positive, β_2 is negative and β_3 is not statistically significant.

Table 2.3 shows the regression results, with coefficients presented in percentages. Ex-ante skewness with interpolated to maturities of 30, 60, and 91 days, is presented in the first two, middle two, and last two columns, respectively, from left to right. In this regression analysis, the benchmark for DSKEW is the change in ex-ante skewness around the management guidance when the analysts' estimations and managers' forecasts are not comparable (i.e., neutral news).

[Insert Table 2.3 here]

Columns (1) and (2) present the regression results when ex-ante skewness represents investors' beliefs regarding stock price crash risk in the future 30 days. In Column (1), the coefficient on *Bad* is 19.93, indicating that ex-ante skewness increases significantly by an additional 0.20 after the disclosure of bad news compared to the benchmark. This suggests that investors' subjective stock price crash risk for the next 30 days decreases more than the benchmark by 0.20. Similarly, the negative coefficient on the *Good* (-10.45) reveals that the decrease in ex-ante skewness following the disclosure of good news is larger in magnitude by 0.10 compared to the benchmark. In Column (2), after adding the control variables, the magnitude of the coefficients decreases. The coefficient on *Neutral* is positive but not statistically significant, indicating that there is no significant difference in the change in ex-ante skewness around disclosures with no surprises compared to the benchmark.

We find consistent results in the last four columns. As the maturity of ex-ante skewness increases, representing a longer forecasting period, the magnitude of the coefficients decreases. This suggests a more stable change in investors' beliefs regarding future stock price crash risk. Specifically, the coefficient on the *Bad* decreases from 19.93 in Column (1) to 7.54 in Column (5) as the expectation period increases from 30 days to 91 days. This change indicates that the decrease in ex-ante crash risk following the disclosure of bad news forecasts from managers becomes less pronounced over longer time horizons. Similarly, the coefficients on the *Good* increases from -10.45 in Column (1) to -3.77 in Column (5) as the expectation period increases from 30 days to 91 days. This result implies that the increase in ex-ante crash risk when good news forecasts are reported becomes relatively smaller over longer timeframes.

In conclusion, we find that investors' subjective stock price crash risk decreases following the disclosure of bad news forecasts from managers and increases when good news forecasts are reported one day after the management guidance disclosure day. These findings hold across different time horizons, indicating a consistent pattern in investors' reactions to different types of guidance disclosures.

2.4.4 Heterogeneity analysis for the change in ex-ante skewness

2.4.4.1 Level of informativeness of management guidance

As the change in ex-ante crash risk in the short window around the management guidance could be attributed to the decrease in information asymmetry, we examine whether the level of informativeness of management guidance affects the change in investors' expectations of short-term stock price crash risk. To proxy the level of informativeness of management guidance, we use the periodicity of management forecasts and whether more than one guidance measure is reported in each management guidance press.

Table 2.4 presents the regression results when the periodicity of forecasts is considered. We define the dummy variable QTR that equals one if the managers guide the quarterly forecasts and zero if the managers guide the annual forecasts. To examine the impacts of forecasts periodicity, we employ the interaction term between disclosure and QTR. Columns (1) and (2) of Table 2.4 show the regression coefficients and corresponding t-statistics when ex-ante skewness captures investors' beliefs regarding stock price crash risk in the future 30 days. The coefficients on the interaction term $Bad \times QTR$ are 31.63 and 24.55, with t-statistics of 6.86 and 5.5, respectively, indicating that the increase in the ex-ante skewness following the bad news disclosures is 0.32 higher for quarterly forecasts than annual forecast. In the economic terms, the coefficient on $Bad \times QTR$ in column (1), 31.63, indicates an additional increase of 33% of standard deviation of $DSKEW$ for quarterly forecasts compared to annual forecasts. We find consistent results when ex-ante skewness captures investors' beliefs regarding stock price crash risk in the future 60 days and 91 days. The coefficients in columns (3) and (5) are 19.13 (t-stat=7.11) and 12.27 (t-stat=6.10), respectively. These results indicate that the increase in ex-ante skewness following bad news disclosures is 0.07 and 0.06 higher for quarterly forecasts compared to annual forecasts, for the next 60 and 91 days, respectively. In the economic term, the coefficients of interaction term $Bad \times QTR$ in columns (3) and (5) correspond to an additional increase of 33% and 28% of standard deviation of $DSKEW$ for quarterly forecasts than annual forecasts, respectively. However, we do not observe statistically significant impacts of quarterly forecasts on the changes in ex-ante skewness following good news disclosures.

[Insert Table 2.4 here]

Table 2.5 presents the regression results when we examine whether multiple guidance measures are reported in each management guidance press. We define $Multi\_type$ as a dummy variable that equals one if there is more than one guidance measures are forecasted and zero otherwise. To test the effect of $Multi\_type$,we employ the interaction term between disclosure and $Multi\_type$. Columns (1) and (2) of Table 2.5 report the regression results when ex-ante skewness captures investors' beliefs regarding stock price crash risk in the future 30 days. The coefficients on the interaction term are 26.2 (t-stat = 4.17) and 15.88 (t-stat = 2.60), respectively. These results indicate that the increase in ex-ante skewness following bad news disclosures is 0.26 higher when there are multiple guidance measures reported compared to when there is only one measure included. In the economic term, the coefficient on interaction term $Bad \times Multi\_type$ in column (1) indicates an additional increase of 27.6% of standard deviation of $DSKEW$ for multiple guidance measures compared to only one measure. We find consistent results when the ex-ante skewness captures investors' beliefs regarding stock price crash risk in the future 60 days and 91 days. The coefficients in columns (3) and (5) are 15.66 (t-stat = 3.96) and 9.88 (t-stat = 2.60), respectively, indicating that the ex-ante skewness increase for next 60 days and 91 days following the bad news disclosures is 0.16 and 0.10 higher when there are multiple guidance measures compared to only one guidance measure. The corresponding economic magnitudes are additional increases of 27.7% and 24% of standard deviation of $DSKEW$. These results suggest that the presence of multiple guidance measures reported in each management guidance press leads to a higher increase in ex-ante skewness following bad news disclosures, indicating a stronger impact on investors' expectations of short-term stock price crash risk.

[Insert Table 2.5 here]

For good news disclosures, the results are significant only for the 60 days expectations. In Columns (3) and (4), when capturing expectations for the next 60 days, the decrease in ex-ante skewness following good news disclosures is statistically significant at 5% (coeff. = -9.14; t-stat = -2.56) and 10% (coeff. = -6.14; t-stat = -1.77) levels, respectively[20]. Additionally, we do not observe significant impacts on the change in ex-ante skewness following neutral news disclosures.

2.4.4.2 Bad news disclosures and information asymmetry

To better understand the mechanisms behind the change in investors' belief in stock price crash risk around the management guidance disclosure, we conduct the heterogeneity analysis based on firms' characteristics and stock price-level variations, following the previous literature for the bad news and good news disclosures.

According to the agency-based model in Jin and Myers (2006), crash risk is linked to firms'

[20] The weak results following good news are consistent with our expectations, as the theoretical model does not directly address changes in ex-ante skewness after good news disclosures. The model primarily focuses on the improvement of information transparency for bad news.

information opacity. This information asymmetry allows managers to hide and absorb bad news. When the accumulation of bad news becomes excessive, firms are forced to reveal the bad news all at once, leading to an extreme stock price decrease. Based on the implications of this theory, investors' expectations of future stock price crash risk should decrease more for firms with high information asymmetry after bad news has been disclosed, as managers from those firms are more likely to hide negative information. We test this prediction using the number of analysts following a firm as a proxy for information asymmetry. Firms with more analysts covering them are more likely to have lower information asymmetry.

[Insert Table 2.6 here]

Table 2.6 reports the results of regressions that examine how the change in investors' belief in future stock price crashes varies cross-sectionally with the number of analysts following ($ACov$), with coefficients presented in percentages. We find that the coefficients on the interaction term $Bad \times ACov$ are negative and significant at the 1% level in all columns, indicating that the increase in ex-ante skewness following the bad news disclosure is smaller for firms with a larger number of analysts covering them. For example, in Column (1) of Table 2.6, the coefficient on interaction term $Bad \times ACov$ is -13.03 (t-stat = 3.37), indicating a smaller increase in ex-ante skewness for firms with low information asymmetry (larger number of analysts). There are no significant results for good news, suggesting alternative explanations for the change in ex-ante skewness following good news disclosures. The decrease in expected future stock price crash risk following the disclosure of bad news could be attributed to a decrease in information asymmetry, which decreases the possibility of bad news accumulation and revelation all at once in the future. We find consistent result when we use the firm size as the alternative proxy for information asymmetry[21].

2.4.4.3 Good news disclosures and litigation risk

The potential explanation for the increase in investors' expectations of future stock price crash risk following good news could be that investors doubt the good news and believe that managers are withholding back bad news. If this explanation holds, we should expect to observe a smaller increase in ex-ante stock price crash risk after the disclosure of good news for firms with less likely to withhold bad news. We use litigation risk as a proxy for the likelihood that managers knowingly withhold bad news. Firms that experience extremely large negative stock returns are more likely to face lawsuits from investors for concealing negative information. It would be costlier for managers to hide bad news if the litigation risk of a firm is higher. Therefore, we predict that the decrease in ex-ante skewness after the disclosure of good news may be mitigated for firms that are less likely to hide bad news due to the higher cost of doing so.

To test this prediction, we estimate firms' litigation risk following the methodology of Kim and Skinner (2012) and Kim, Wang and Zhang (2019).

$$\begin{aligned} Litigation_{i,t} = &-7.883 + 0.556 \times FSP_{i,t} + 0.518 \times LnAssets_{i,t-1} \\ &+ 0.982 \times Sales\ Growth_{i,t-1} + 0.379 \times Return_{i,t-1} \\ &- 0.108 \times Return\ Skewness_{i,t-1} + 25.635 \times Rsturn\ Std\ Dev_{i,t-1} \\ &+ 7 \times 10^{-7} \times Turnover_{i,t-1}, \end{aligned} \tag{2.7}$$

where FSP is a dummy variable that equals one if the firm is in the biotech (SIC codes 2833–2836 and 8731–8734), computer (3570–3577 and 7370–7374), electronics (3600–3674), or retail (5200–5961) industries, and 0 otherwise. $LnASssets$ is the natural log of total assets at the end of the year. $Sales\ Growth$ is the change in sales scaled by total assets. $Return$ is the market-adjusted

[21] See Appendix B.

12-month accumulative stock returns. $Return\ Skewness$ and $Return\ Std\ Dev$ are skewness and standard deviation of firms' a 12-month returns. $TURNOVER$ is the accumulative trading volume over 12-month period scaled by the total shares outstanding at the beginning of the year. $Litgition$ is estimated at the stock-year level. A high value of $Litigation_{i,t}$ represents high litigation risk for stock i in year t.

[Insert Table 2.7 here]

Table 2.7 shows the regression results for good news disclosures and litigation risk. The coefficients of the interaction term between $Good$ and $Litigation$ are positive in general. Importantly, we observe that the coefficients on the interaction term are statistically significant at the 5% level in column (5) and (6). The result suggests that when a firm discloses good news in its earnings guidance, the positive impact on investors' expected stock price crash risk in the future 91 days is mitigated if the firm's litigation risk is higher. However, the mitigation effect of litigation risk is less statistically significant on the positive impact of guided good news on investors' expected stock price crash risk within shorter horizons. One possible explanation is that the litigation risk, estimated annually, reflects longer-term forecasts regarding the possibility of firms being sued by investors for extreme negative stock returns in the future. These results are consistent with the explanation that investors are suspicious that managers are knowingly hiding bad news, especially for firms where managers have a higher likelihood of manipulating voluntary information disclosure for their short-term interests with lower cost.

2.4.5 Ex-post stock price crash risk predictions

In this section, we investigate whether the change in investors' expectations of future stock price crash risk around the disclosure of management guidance can predict realized stock price crash in the following month. Since investors' belief on the change in future stock price crash risk is opposite following good versus bad news, we examine the predictability of future realized stock price crashes with an interaction term using the the following regression model:

$$\begin{aligned} rSKEW_{i,t+3,t+33} &= \alpha + \beta_1 \times DSKEW_{i,t,t+N} + \beta_2 \times stdSKEW_{i,t-3,t-3+N} \\ &\quad +\beta_3 \times rSKEW_{i,t-31,t-1} + \boldsymbol{\beta} \times \boldsymbol{X} + FE_i + FE_{qtr} \end{aligned} \tag{2.8}$$

where the main independent variable is $DSKEW$, which is the change in ex-ante skewness estimated using Eq. (2.4), capturing the change in ex-ante skewness from day $t-2$ to $t+2$. $DSKEW_{i,t,t+N}$ measures the change in ex-ante skewness of stock i around the disclosure around day t, estimated using the option with N days maturity. The dependent variable is the realized 30-day skewness that is estimated using daily stock returns starting from one day after the latest ex-ante skewness used in $DSKEW$ estimation at day $t+2$. Since our focus is on predicting the change in ex-ante skewness around the disclosure day, we control the level of past ex-ante skewness measured on day $t=-3$ ($stdSKEW_{i,t-3,t-3+N}$). Additionally, we control for the past one-month realized skewness ($rSKEW_{i,t-31,t-1}$), which is estimated using daily realized stock.

The results are reported in Columns (1) and (2) of Table 2.8. The coefficient on $DSKEW$ is positive and statistically significant at the 10% level in Column (1). The result suggests that $DSKEW$ has predictive power for future realized stock price crash risk following the management guidance. The result is not affected when the control variables are added as shown in column (2). We do not find statistically significant results when the change in investors' belief about stock price crash for the future 60 days and 91 days, which are shown in Appendix C. In summary, we find that the increase in investors' expectations following the management guidance can significantly predict short-term realized stock price crash risk. However, we do not observe significant predictability for long-term realized skewness.

[Insert Table 2.8 here]

2.5 Conclusion

In this chapter, we investigate whether the disclosure of management guidance can influence investors' beliefs regarding future stock price crash risk, focusing on a narrow window around the disclosure day. We use option-implied ex-ante skewness as a measure of investors' expectations of future stock price crash risk. Consistent with previous theoretical models, our findings suggest that investors' perceived stock price crash risk decreases when firms' guidance is below the mean of analysts' estimation (bad news) and increases when firms' guidance is above the mean of analysts' estimation (good news). As the placebo test, we find there is no significant change in perceived stock price crash risk following the neutral news, when the mean of analysts' estimations falls within the expected range reported by the managers.

We further demonstrate that the increase in ex-ante skewness (reflecting the decrease in subjective stock price crash risk) following bad news is affected by the level of informativeness of management guidance. The decreases in expected stock price crash risk following bad news can be attributed to a reduction in information asymmetry related to bad news. When bad news is disclosed, it becomes less likely that managers would knowingly conceal this information, reducing the likelihood of a severe stock price drop when accumulated bad news is eventually revealed in the future. Regarding the increase in expected stock price crash risk after good news disclosure, our finding supports the argument that investors may be suspicious about the good news announced by managers to conceal the actual bad news for the managers' own interests. Our empirical results also provide evidence that changes in expected stock price crash risk around management guidance can predict realized stock price crash risk in the future 30 days following the management forecasts.

References

An, Li, 2016. Asset pricing when traders sell extreme winners and losers. *Review of Financial Studies* 29 (3), 823-861.

Ang, Andrew, Robert J Hodrick, Yuhang Xing, and Xiaoyan Zhang, 2006. The cross-section of volatility and expected returns. *Journal of Finance* 61 (1), 259-299.

Baginski, Stephen P, Edward J Conrad, and John M Hassell, 1993. The effects of management forecast precision on equity pricing and on the assessment of earnings uncertainty. *Accounting Review* 913-927.

Baker, Andrew C, David F Larcker, and Charles CY Wang, 2022. How much should we trust staggered difference-in-differences estimates? *Journal of Financial Economics* 144 (2), 370-395.

Barber, Brad M, and Terrance Odean, 2000. Trading is hazardous to your wealth: The common stock investment performance of individual investors. *Journal of Finance* 55 (2), 773-806.

Barber, Brad M., and Terrance Odean, 2015. Online investors: Do the slow die first? *Review of Financial Studies* 15 (2), 455-488.

Barry, Christopher B., and Stephen J. Brown, 1985. Differential information and security market equilibrium. *Journal of Financial and Quantitative Analysis* 20 (4), 407-422.

Ben-David, Itzhak, and David Hirshleifer, 2012. Are investors really reluctant to realize their losses? Trading responses to past returns and the disposition effect. *Review of Financial Studies* 25 (8), 2485-2532.

Bergman, Nittai K, and Sugata Roychowdhury, 2008. Investor sentiment and corporate disclosure. *Journal of Accounting Research* 46 (5), 1057-1083.

Beyer, Anne, Daniel A Cohen, Thomas Z Lys, and Beverly R Walther, 2010. The financial reporting environment: Review of the recent literature. *Journal of Accounting and Economics* 50 (2-3), 296-343.

Biais, Bruno, Denis Hilton, Karine Mazurier, and Sébastien Pouget, 2005. Judgemental overconfidence, self-monitoring, and trading performance in an experimental financial market. *Review of Economic Studies* 72 (2), 287-312.

Blankespoor, Elizabeth, Ed deHaan, and Ivan Marinovic, 2020. Disclosure processing costs, investors' information choice, and equity market outcomes: A review. *Journal of Accounting and Economics* 101344.

Bleck, Alexander, and Xuewen Liu, 2007. Market transparency and the accounting regime. *Journal of Accounting Research* 45 (2), 229-256.

Cengiz, Doruk, Arindrajit Dube, Attila Lindner, and Ben Zipperer, 2019. The effect of minimum wages on low-wage jobs. *Quarterly Journal of Economics* 134 (3), 1405-1454.

Chang, Yen-Cheng, Alexander Ljungqvist, and Kevin Tseng, 2023. Do corporate disclosures constrain strategic analyst behavior? *Review of Financial Studies* 36 (8), 3163-3212.

Chang, Yen-Cheng, Pei-Jie Hsiao, Alexander Ljungqvist, and Kevin Tseng, 2022. Testing disagreement models. *Journal of Finance* 77 (4), 2239-2285.

Chen, Joseph, Harrison Hong, and Jeremy C. Stein, 2001. Forecasting crashes: Trading volume, past returns, and conditional skewness in stock prices. *Journal of Financial Economics* 61 (3), 345-381.

Chen, Shuping, Dawn Matsumoto, and Shiva Rajgopal, 2011. Is silence golden? An empirical analysis of firms that stop giving quarterly earnings guidance. *Journal of Accounting and Economics* 51 (1), 134-150.

Cheng, Qiang, and Kin Lo, 2006. Insider trading and voluntary disclosures. *Journal of Accounting Research* 44 (5), 815-848.

Chui, Andy CW, Sheridan Titman, and KC John Wei, 2010. Individualism and momentum around the world. *Journal of Finance* 65 (1), 361-392.

Coller, Maribeth, and Teri Lombardi Yohn, 1997. Management forecasts and information asymmetry: An examination of bid-ask spreads. *Journal of Accounting Research* 35 (2), 181-191.

Cotter, Julie, Irem Tuna, and Peter D Wysocki, 2006. Expectations management and beatable targets: How do analysts react to explicit earnings guidance? *Contemporary Accounting Research* 23 (3), 593-624.

Daniel, Kent , David Hirshleifer, and Avanidhar Subrahmanyam, 2001. Overconfidence, arbitrage, and equilibrium asset pricing. *Journal of Finance* 56 (3), 921-965.

Daniel, Kent, David Hirshleifer, and Avanidhar Subrahmanyam, 1998. Investor psychology and security market under-and overreactions. *Journal of Finance* 53 (6), 1839-1885.

De Bondt, Werner FM, and Richard H Thaler, 1995. Financial decision-making in markets and firms: A behavioral perspective. *Handbooks in operations research and management science* 9 385-410.

DeFond, Mark L, Mingyi Hung, Siqi Li, and Yinghua Li, 2015. Does mandatory ifrs adoption affect crash risk? *Accounting Review* 90 (1), 265-299.

Ertugrul, Mine, Jin Lei, Jiaping Qiu, and Chi Wan, 2017. Annual report readability, tone ambiguity, and the cost of borrowing. *Journal of Financial and Quantitative Analysis* 52 (2), 811-836.

Fama, E. F., and K. R. French, 1992. The cross-section of expected stock returns. *Journal of Finance* 47 (2), 427-465.

Fama, E. F., and K. R. French, 1993. Common risk factors in the returns on stocks and bonds. *Journal of Financial Economics* 33 (1), 3-56.

Fischhoff, Baruch, Paul Slovic, and Sarah Lichtenstein, 1977. Knowing with certainty: The appropriateness of extreme confidence. *Journal of Experimental Psychology: Human perception and performance* 3 (4), 552.

Gao, Meng, and Jiekun Huang, 2020. Informing the market: The effect of modern information technologies on information production. *Review of Financial Studies* 33 (4), 1367-1411.

Gao, Pingyang, and Pierre Jinghong Liang, 2013. Informational feedback, adverse selection, and optimal disclosure policy. *Journal of Accounting Research* 51 (5), 1133-1158.

Gao, Xiaohui, and Jay R. Ritter, 2010. The marketing of seasoned equity offerings. *Journal of Financial Economics* 97 (1), 33-52.

Gervais, S., and T. Odean, 2001. Learning to be overconfident. *Review of Financial Studies* 14 (1), 1-27.

Glaser, Markus, and Martin Weber, 2007. Overconfidence and trading volume. *The Geneva Risk and Insurance Review* 32 (1), 1-36.

Goldstein, Itay, Shijie Yang, and Luo Zuo, 2023. The real effects of modern information technologies: Evidence from the edgar implementation. *Journal of Accounting Research*.

Grinblatt, Mark, and Bing Han, 2005. Prospect theory, mental accounting, and momentum. *Journal of Financial Economics* 78 (2), 311-339.

Hermalin, Benjamin E, and Michael S Weisbach, 2012. Information disclosure and corporate governance. *Journal of Finance* 67 (1), 195-233.

Hsu, Charles, Rencheng Wang, and Benjamin C Whipple, 2022. Non-gaap earnings and stock price crash risk. *Journal of Accounting and Economics* 73 (2-3), 101473.

Huang, Zeqiong, Joseph D Piotroski, and Chloe L Xie, 2023. Strategic disclosure and investor loss aversion *Working Paper*.

Hutton, Amy P, Gregory S Miller, and Douglas J Skinner, 2003. The role of supplementary statements with management earnings forecasts. *Journal of Accounting Research* 41 (5), 867-890.

Hutton, Amy P., Alan J. Marcus, and Hassan Tehranian, 2009. Opaque financial reports, r2, and crash risk. *Journal of Financial Economics* 94 (1), 67-86.

Isakov, Dušan, and Christophe Pérignon, 2001. Evolution of market uncertainty around earnings announcements. *Journal of Banking and Finance* 25 (9), 1769-1788.

Jin, Li, and Stewart C. Myers, 2006. R2 around the world: New theory and new tests. *Journal of Financial Economics* 79 (2), 257-292.

Kahneman, Daniel, 2011. *Thinking, fast and slow* (Farrar, Straus and Giroux, NY).

Kim, Chansog, Ke Wang, and Liandong Zhang, 2019. Readability of 10-k reports and stock price crash risk. *Contemporary Accounting Research* 36 (2), 1184-1216.

Kim, Jeong-Bon, Leye Li, Louise Yi Lu, and Yangxin Yu, 2016. Financial statement comparability and expected crash risk. *Journal of Accounting and Economics* 61 (2), 294-312.

Kim, Jeong-Bon, and Liandong Zhang, 2016. Accounting conservatism and stock price crash risk: Firm-level evidence. *Contemporary Accounting Research* 33 (1), 412-441.

Kim, Jeong‐Bon, and Liandong Zhang, 2014. Financial reporting opacity and expected crash risk: Evidence from implied volatility smirks. *Contemporary Accounting Research* 31 (3), 851-875.

Kim, Yong Hyuck, Zoran Ivkovich, and Dmitriy Muravyev, 2021. Causal effect of information costs on asset pricing anomalies. Working paper, available at SSRN 3921785.

Kothari, Sabino P, Susan Shu, and Peter D Wysocki, 2009. Do managers withhold bad news? *Journal of Accounting Research* 47 (1), 241-276.

Kumar, Alok, 2009. Hard-to-value stocks, behavioral biases, and informed trading. *Journal of Financial and Quantitative Analysis* 44 (6), 1375-1401.

Leary, Mark T., and Michael R. Roberts, 2010. The pecking order, debt capacity, and information asymmetry. *Journal of Financial Economics* 95 (3), 332-355.

Li, Si, and Xintong Zhan, 2019. Product market threats and stock crash risk. *Management Science* 65 (9), 4011-4031.

Liu, Yibin, 2019. Information acquisition costs and misreporting: Evidence from the implementation of edgar. Working paper.

Lu, Yao, and Douglas J Skinner, 2020. Moving forward: Management guidance and earnings announcement returns. *Working Paper*.

Maskara, Pankaj K., and Donald J. Mullineaux, 2011. Information asymmetry and self-selection bias in bank loan announcement studies. *Journal of Financial Economics* 101 (3), 684-694.

Miller, Gregory S, 2002. Earnings performance and discretionary disclosure. *Journal of Accounting Research* 40 (1), 173-204.

Noe, Christopher F., 1999. Voluntary disclosures and insider transactions. *Journal of Accounting and Economics* 27 (3), 305-326.

Odean, Terrance, 1998. Are investors reluctant to realize their losses? *Journal of Finance* 53 (5), 1775-1798.

Odean, Terrance, 1998. Volume, volatility, price, and profit when all traders are above average. *Journal of Finance* 53 (6), 1887-1934.

Odean, Terrance, 1999. Do investors trade too much? *American Economic Review* 89 (5), 1279-

1298.
Patell, James M, and Mark A Wolfson, 1979. Anticipated information releases reflected in call option prices. *Journal of Accounting and Economics* 1 (2), 117-140.
Patell, James M, and Mark A Wolfson, 1981. The ex ante and ex post price effects of quarterly earnings announcements reflected in option and stock prices. *Journal of Accounting Research* 434-458.
Patell, Jm, 1976. Corporate forecasts of earnings per share and stock-price behavior-empirical tests. *Journal of Accounting Research* 14 246-276.
Penman, Stephen H, 1980. An empirical investigation of the voluntary disclosure of corporate earnings forecasts. *Journal of Accounting Research* 132-160.
Rogers, Jonathan L, Douglas J Skinner, and Andrew Van Buskirk, 2009. Earnings guidance and market uncertainty. *Journal of Accounting and Economics* 48 (1), 90-109.
Rogers, Jonathan L, and Phillip C Stocken, 2005. Credibility of management forecasts. *Accounting Review* 80 (4), 1233-1260.
Rogers, Jonathan L, and Andrew Van Buskirk, 2013. Bundled forecasts in empirical accounting research. *Journal of Accounting and Economics* 55 (1), 43-65.
Scheinkman, Jose A, and Wei Xiong, 2003. Overconfidence and speculative bubbles. *Journal of Political Economy* 111 (6), 1183-1220.
Schneider, Paul, Christian Wagner, and Josef Zechner, 2020. Low-risk anomalies? *Journal of Finance* 75 (5), 2673-2718.
Singh, Dharmendra, Garima Malik, and Aruna Jha, 2024. Overconfidence bias among retail investors: A systematic review and future research directions. *Investment Management and Financial Innovations* 21 (1), 302-316.
Skinner, Douglas J., 1994. Why firms voluntarily disclose bad news. *Journal of Accounting Research* 32 (1), 38-60.
Skinner, Douglas J., 1997. Earnings disclosures and stockholder lawsuits. *Journal of Accounting and Economics* 23 (3), 249-282.
Statman, M., S. Thorley, and K. Vorkink, 2006. Investor overconfidence and trading volume. *Review of Financial Studies* 19 (4), 1531-1565.
Zhang, X Frank, 2006. Information uncertainty and stock returns. *Journal of Finance* 61 (1), 105-137.

Figures and Tables

Table 1.1 The timetable of the EDGAR system implementation
This table shows the EDGAR implementation timelines. U.S. public firms are assigned by the SEC into ten groups to join the EDGAR platform mandatorily from 1993 to 1996. The implementation date is the first day when the firms in the corresponding groups begin to join EDGAR gradually. We assume that firms finish joining the EDGAR platform in the quarter after the implementation date (implementation year-quarter) and investors could retrieve the financial information two quarters later (effective year-quarter).

Implementation group	Implementation date	Implementation year-quarter	Effective year-quarter
1	April 26 1993	1993Q2	1993Q4
2	July 19 1993	1993Q3	1994Q1
3	October 4 1993	1993Q4	1994Q2
4	December 6 1993	1993Q4	1994Q2
5	August 1994	1994Q3	1995Q1
6	November 1994	1994Q4	1995Q2
7	May 1995	1995Q2	1995Q4
8	August 1995	1995Q3	1996Q1
9	November 1995	1995Q4	1996Q2
10	May 1996	1996Q2	1996Q4

Table 1.2 Summary statistics

This table reports the summary statistics of the variables used in the empirical analysis, including mean (Mean), standard deviation (Std Dev), and quartiles (P25, P50, and P75). Trading volume (*VO*) is calculated as the natural logarithm of one plus the sum of shares traded (in thousands) in a quarter for stocks that are traded at least once by retail investors in our sample. Dollar Volume (*DVO*) is calculated as the natural logarithm of one plus the sum of dollar shares traded (in thousands) in a quarter. Trading Frequency (*Freq*) is calculated as the natural logarithm of one plus sum of number of trades for each stock in a quarter. Post-trade sell-buy return differential (*PTSBD*) is the mean difference between 30-day return following all purchases in each stock in a quarter and 30-day return following all sells in the same stock in the same quarter. The 30-day return is calculated as the dollar amount times excess cumulative returns in 30 days following the trades. *PTSBD* is winsorized at 1 and 99 percentile levels. *Post* equals one if the quarter is in or after the effective year-quarter of the EDGAR implementation for a firm and is zero otherwise. *Market beta* is estimated using 60-month rolling-window regressions of excess returns on the market returns for each stock in each month. Firm size (*Size*) is the natural logarithm of the market value of the equity, calculated in million dollars. Momentum (*MOM*) is the natural logarithm of one plus the cumulative return of a stock in the last year excluding the most recent month. Book-to-market equity (*B/M*) is computed as the natural logarithm of book-to-market equity. *Price* is the end-of-month stock price. *Turnover* is measured as the daily average of the number of shares traded in a quarter divided by the number of shares outstanding, in percentage. The number of shares traded for Nasdaq stocks is divided by 2. *Dividend dummy* is one if the stock pays dividends at least once during the past fiscal year. *Nasdaq dummy* variable is one if the stock is listed on the Nasdaq exchange. Institutional ownership (*INST*) is calculated as the natural logarithm of one plus the institutional ownership scaled by shares outstanding for each stock in each quarter. Analyst coverage (*ACov*) is calculated as the natural logarithm of one plus the average number of analyst coverage for each stock in each quarter. *Bid-ask spread* is the average bid-ask spread in percentage for each stock in each month. Idiosyncratic volatility (*IVOL*) is the standard deviation of the daily residuals from the Fama-French three-factor model. Firm age (*Age*) is measured as the number of years since the firm was first covered by the CRSP. The sample includes ordinary common stocks of U.S. companies listed on major exchanges (NYSE, Amex, and Nasdaq). The sample period is from January 1991 to December 1996.

	Mean	Std Dev	P25	P50	P75	N
Trading Volume (VO)	0.84	1.12	0.00	0.24	1.44	34,212
Dollar Volume (DVO)	2.09	2.26	0.00	1.57	3.93	34,212
Trading Frequency (*Freq*)	1.16	1.32	0.00	0.69	2.08	34,212
PTSBD	-12.30	703.03	-226.72	11.22	202.08	18,731
Post	0.36	0.48	0.00	0.00	1.00	34,212
Market beta	1.07	0.79	0.60	1.01	1.46	34,212
Size	5.01	1.88	3.64	4.88	6.22	34,212
MOM	0.09	0.42	-0.12	0.10	0.31	34,212
B/M	-0.58	0.86	-1.05	-0.50	-0.05	34,212
Price	19.12	19.19	6.63	14.50	26.25	34,212
Turnover	0.27	0.32	0.09	0.18	0.33	34,212
Dividend dummy	0.51	0.50	0.00	1.00	1.00	34,212
Nasdaq dummy	0.55	0.50	0.00	1.00	1.00	34,212
INST	0.09	0.16	0.00	0.00	0.11	34,212
ACov	0.84	0.71	0.00	0.81	1.36	34,212
Bid-ask spread	4.26	4.25	1.59	2.88	5.26	34,212
IVOL	2.93	2.37	1.43	2.29	3.63	34,212
Age	15.98	13.74	6.41	11.54	21.89	34,212

Table 1.3 Effect of EDGAR implementation on the overconfidence (staggered diff-in-diff)
The table reports the coefficients and t-statistics from regressions of the overconfidence on the dummy variable *Post* from Eq. (1.3). Dependent variables are three trading measures to overconfidence. Trading volume (*VO*) is calculated as the natural logarithm of one plus the sum of shares traded (in thousands) in a quarter for stocks that are traded at least once by retail investors in our sample. Dollar Volume (*DVO*) is calculated as the natural logarithm of one plus the sum of dollar shares traded (in thousands) in a quarter. Trading Frequency (*Freq*) is calculated as the natural logarithm of one plus sum of number of trades for each stock in a quarter. *Post* is an indicator that equals one if quarter t is on or after the effective date of each implementation phase and equals zero if quarter t is before that date. The definitions of other explanatory variables are described in Table 1.2. The sample includes ordinary common stocks of U.S. companies listed on major exchanges (NYSE, Amex, and Nasdaq). The sample period is from January 1991 to December 1996. Standard errors are clustered at the stock and year-quarter levels. The t-statistics are reported in parentheses. ***, **, and *indicate statistical significance at the 1%, 5%, and 10% levels, respectively.

	Trading volume		Dollar volume		Trading frequency	
	(1)	(2)	(3)	(4)	(5)	(6)
Post	-0.087***	-0.097***	-0.125***	-0.149***	-0.101***	-0.115***
	(-4.35)	(-5.18)	(-3.76)	(-4.87)	(-4.23)	(-5.09)
Market beta		0.026		0.057*		0.018
		(1.37)		(1.99)		(1.04)
Size		0.144***		0.493***		0.267***
		(4.39)		(11.02)		(8.96)
MOM		-0.004		0.108***		0.036
		(-0.18)		(3.31)		(1.64)
B/M		-0.007		-0.038		-0.026
		(-0.42)		(-1.34)		(-1.47)
Price		-0.007***		-0.003*		-0.003**
		(-3.05)		(-2.10)		(-2.55)
Turnover		0.388***		0.594***		0.393***
		(3.12)		(3.031)		(3.11)
Dividend dummy		0.010		-0.020		-0.010
		(0.35)		(-0.46)		(-0.44)
Nasdaq dummy		-0.103*		-0.209*		-0.131**
		(-1.92)		(-1.91)		(-2.19)
INST		-0.076		0.013		-0.071
		(-0.87)		(0.09)		(-0.73)
ACov		0.019		0.044		0.029
		(0.92)		(1.22)		(1.44)
Bid-ask spread		-0.002		0.004		0.002
		(-0.52)		(0.82)		(0.61)
IVOL		0.013***		0.018**		0.013**
		(3.12)		(2.40)		(2.68)
Firm FE	Yes	Yes	Yes	Yes	Yes	Yes
Year-quarter FE	Yes	Yes	Yes	Yes	Yes	Yes
Observations	34,194	34,194	34,194	34,194	34,194	34,194
Adj R-squared	0.746	0.755	0.800	0.813	0.815	0.827

Table 1.4 Effect of EDGAR implementation on the overconfidence (stacked diff-in-diff)

The table reports the coefficients and *t*-statistics from regressions of the overconfidence on the dummy variable *Post* from Eq. (1.4). Dependent variables are three trading measures to overconfidence. Trading volume (*VO*) is calculated as the natural logarithm of one plus the sum of shares traded (in thousands) in a quarter for stocks that are traded at least once by retail investors in our sample. Dollar Volume (*DVO*) is calculated as the natural logarithm of one plus the sum of dollar shares traded (in thousands) in a quarter. Trading Frequency (*Freq*) is calculated as the natural logarithm of one plus sum of number of trades for each stock in a quarter. *Post* is an indicator that equals one if quarter t is on or after the effective date of each implementation phase and equals zero if quarter t is before that date. The definitions of other explanatory variables are described in Table 1.2. The sample includes ordinary common stocks of U.S. companies listed on major exchanges (NYSE, Amex, and Nasdaq). The sample period is from January 1991 to December 1996. Standard errors are clustered at the stock and year-quarter levels. The *t*-statistics are reported in parentheses. ***, **, and *indicate statistical significance at the 1%, 5%, and 10% levels, respectively.

	Trading volume		Dollar volume		Trading frequency	
	(1)	(2)	(3)	(4)	(5)	(6)
Post	-0.144***	-0.138***	-0.192***	-0.206***	-0.131***	-0.135***
	(-4.60)	(-4.80)	(-3.49)	(-4.38)	(-3.83)	(-4.27)
Market beta		0.030		0.063*		0.020
		(1.25)		(1.76)		(0.92)
Size		0.196***		0.520***		0.292***
		(4.50)		(5.93)		(5.49)
MOM		-0.017		0.062		0.015
		(-0.69)		(1.48)		(0.53)
B/M		0.019		0.014		0.001
		(0.92)		(0.36)		(0.02)
Price		-0.012***		-0.006		-0.006*
		(-5.21)		(-1.18)		(-1.89)
Turnover		0.263**		0.377**		0.268**
		(2.54)		(2.34)		(2.57)
Dividend dummy		0.002		-0.035		-0.011
		(0.06)		(-0.64)		(-0.30)
Nasdaq dummy		-0.121		-0.276		-0.139
		(-1.20)		(-1.38)		(-1.26)
INST		0.074		0.174		0.127
		(0.81)		(0.95)		(1.14)
ACov		0.034		0.071*		0.050**
		(1.38)		(1.80)		(2.16)
Bid-ask spread		-0.001		0.001		0.001
		(-0.26)		(0.11)		(0.13)
IVOL		0.014**		0.018		0.014*
		(2.16)		(1.67)		(2.10)
Controls	No	Yes	No	Yes	No	Yes
Firm FE	Yes	Yes	Yes	Yes	Yes	Yes
Year-quarter FE	Yes	Yes	Yes	Yes	Yes	Yes
Observations	26,234	26,234	26,234	26,234	26,234	26,234
Adj R-squared	0.750	0.757	0.802	0.811	0.815	0.823

Table 1.5 Effect of EDGAR implementation on the overconfidence: Young versus old firms

The table reports the coefficients and *t*-statistics from the regressions of the trading level on the dummy variable *Post*. We run the regression for young firms and old firms separately. Young firms (Old firms) are defined as firms whose age is lower than or equal to (higher than) the median firm age in our sample. Firm age is measured as the number of years since the firm was first covered by the CRSP. Panel A shows the regression results when the dependent variable is trading volume (*VO*), which is defined as the natural logarithm of the sum of shares traded (in thousands) of stock i in quarter t by retail investors in the sample. Panel B shows the regression results when the dependent variable is Dollar Volume (*DVO*), which is calculated as the natural logarithm of one plus the sum of dollar shares traded (in thousands) in a quarter. Panel C shows the regression results when the dependent variable is the Trading Frequency (*Freq*), which is calculated as the natural logarithm of one plus sum of number of trades for each stock in a quarter. *Post* is an indicator that equals one if quarter t is on or after the effective date of each implementation phase and zero if quarter t is before that date. Control variables are the same as those in Table 1.2. Columns (1), (2), (5), and (6) are baseline staggered diff-in-diff OLS regressions, whereas columns (3), (4), (7), and (8) are stacked diff-in-diff regressions. The sample period is from January 1991 to December 1996. Standard errors are clustered at the stock and year-quarter levels. The *t*-statistics are reported in parentheses. ***, **, and *indicate statistical significance at the 1%, 5%, and 10% levels, respectively.

Panel A: Dependent variable =	Trading volume							
	Young firms				Old firms			
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
Post	-0.114***	-0.126***	-0.125**	-0.116**	-0.037	-0.047*	-0.059	-0.056
	(-4.05)	(-4.99)	(-2.62)	(-2.73)	(-1.43)	(-1.88)	(-1.47)	(-1.55)
Controls	No	Yes	No	Yes	No	Yes	No	Yes
Firm FE	Yes	Yes			Yes	Yes		
Year-quarter FE	Yes	Yes			Yes	Yes		
Firm FE × Group FE			Yes	Yes			Yes	Yes
Year-quarter FE × Group FE			Yes	Yes			Yes	Yes
Observations	17,083	17,083	12,942	12,942	17,094	17,094	13,148	13,148
Adj R-squared	0.746	0.757	0.728	0.736	0.748	0.759	0.773	0.780
Diff between column (n) and (n+4)	-0.078	-0.079	-0.067	-0.060				
p-value	0.00	0.00	0.00	0.00				

Panel B: Dependent variable =	Dollar volume							
	Young firms				Old firms			
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
Post	-0.138**	-0.189***	-0.166*	-0.184**	-0.083	-0.088*	-0.145*	-0.147**
	(-2.91)	(-4.67)	(-2.02)	(-2.72)	(-1.69)	(-1.85)	(-2.12)	(-2.23)
Controls	No	Yes	No	Yes	No	Yes	No	Yes
Firm FE	Yes	Yes			Yes	Yes		
Year-quarter FE	Yes	Yes			Yes	Yes		
Firm FE × Group FE			Yes	Yes			Yes	Yes

Year-quarter FE × Group FE			Yes	Yes			Yes	Yes
Observations	17,083	17,083	12,942	12,942	17,094	17,094	13,148	13,148
Adj R-squared	0.800	0.817	0.797	0.806	0.800	0.809	0.809	0.817
Diff between column (n) and (n+4)	-0.054	-0.101	-0.021	-0.036				
p-value	0.01	0.00	0.22	0.09				

Panel C: Dependent variable =	Trading frequency							
	Young firms				Old firms			
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
Post	-0.130***	-0.156***	-0.121**	-0.126**	-0.049	-0.057*	-0.068	-0.070
	(-3.96)	(-5.19)	(-2.41)	(-2.94)	(-1.63)	(-1.90)	(-1.29)	(-1.41)
Controls	No	Yes	No	Yes	No	Yes	No	Yes
Firm FE	Yes	Yes			Yes	Yes		
Year-quarter FE	Yes	Yes			Yes	Yes		
Firm FE × Group FE			Yes	Yes			Yes	Yes
Year-quarter FE × Group FE			Yes	Yes			Yes	Yes
Observations	17,083	17,083	12,942	12,942	17,094	17,094	13,148	13,148
Adj R-squared	0.810	0.825	0.805	0.815	0.821	0.831	0.827	0.834
Diff between column (n) and (n+4)	-0.081	-0.099	-0.052	-0.056				
p-value	0.00	0.00	0.00	0.00				

Table 1.6 Effect of EDGAR implementation on the overconfidence: Growth versus value firms

The table reports the coefficients and *t*-statistics from the regressions of the trading level on the dummy variable *Post*. We run the regression for growth firms and value firms separately. Growth firms (Value firms) are defined as firms whose firm's book-to-market ratio is lower than or equal to (higher than) the median in our sample. Panel A shows the regression results when the dependent variable is trading volume (*VO*), which is defined as the natural logarithm of the sum of shares traded (in thousands) of stock i in quarter t by retail investors in the sample. Panel B shows the regression results when the dependent variable is Dollar Volume (*DVO*), which is calculated as the natural logarithm of one plus the sum of dollar shares traded (in thousands) in a quarter. Panel C shows the regression results when the dependent variable is the Trading Frequency (*Freq*), which is calculated as the natural logarithm of one plus sum of number of trades for each stock in a quarter. *Post* is an indicator that equals one if quarter t is on or after the effective date of each implementation phase, and zero if quarter t is before that date. The control variables are the same as those in Table 1.2. Columns (1), (2), (5), and (6) are baseline staggered diff-in-diff OLS regressions, whereas columns (3), (4), (7), and (8) are stacked diff-in-diff regressions. The sample period is from January 1991 to December 1996. Standard errors are clustered at the stock and year-quarter levels. The *t*-statistics are reported in parentheses. ***, **, and *indicate statistical significance at the 1%, 5%, and 10% levels, respectively.

Panel A: Dependent variable =	Trading volume							
	Growth firms				Value firms			
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
Post	-0.091***	-0.109***	-0.177***	-0.172***	-0.069***	-0.068***	-0.056*	-0.063*
	(-3.05)	(-4.07)	(-3.61)	(-3.99)	(-3.45)	(-3.58)	(-1.84)	(-2.13)
Controls	No	Yes	No	Yes	No	Yes	No	Yes
Firm FE	Yes	Yes			Yes	Yes		
Year-quarter FE	Yes	Yes			Yes	Yes		
Firm FE × Group FE			Yes	Yes			Yes	Yes
Year-quarter FE × Group FE			Yes	Yes			Yes	Yes
Observations	17,090	17,090	12,948	12,948	17,079	17,079	13,022	13,022
Adj R-squared	0.754	0.766	0.752	0.760	0.740	0.745	0.744	0.747
Diff between column (n) and (n+4)	-0.023	-0.041	-0.121	-0.109				
p-value	0.04	0.00	0.00	0.00				

Panel B: Dependent variable =	Dollar volume							
	Growth firms				Value firms			
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
Post	-0.131**	-0.172***	-0.268***	-0.263***	-0.115***	-0.107***	-0.068	-0.104*
	(-2.53)	(-3.83)	(-3.08)	(-3.78)	(-3.10)	(-3.04)	(-1.24)	(-1.99)
Controls	No	Yes	No	Yes	No	Yes	No	Yes
Firm FE	Yes	Yes			Yes	Yes		

Year-quarter FE	Yes	Yes			Yes	Yes		
Firm FE × Group FE			Yes	Yes			Yes	Yes
Year-quarter FE × Group FE			Yes	Yes			Yes	Yes
Observations	17,090	17,090	12,948	12,948	17,079	17,079	13,022	13,022
Adj R-squared	0.803	0.818	0.798	0.808	0.790	0.798	0.786	0.791
Diff between column (n) and (n+4)	-0.016	-0.065	-0.200	-0.159				
p-value	0.25	0.00	0.00	0.00				

Panel C: Dependent variable =	Trading frequency							
	Growth firms				Value firms			
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
Post	-0.122***	-0.144***	-0.184***	-0.177***	-0.072***	-0.070***	-0.033	-0.049
	(-3.59)	(-4.88)	(-3.57)	(-4.26)	(-2.96)	(-3.00)	(-0.93)	(-1.34)
Controls	No	Yes	No	Yes	No	Yes	No	Yes
Firm FE	Yes	Yes			Yes	Yes		
Year-quarter FE	Yes	Yes			Yes	Yes		
Firm FE × Group FE			Yes	Yes			Yes	Yes
Year-quarter FE × Group FE			Yes	Yes			Yes	Yes
Observations	17,090	17,090	12,948	12,948	17,079	17,079	13,022	13,022
Adj R-squared	0.819	0.834	0.813	0.823	0.807	0.813	0.805	0.809
Diff between column (n) and (n+4)	-0.050	-0.074	-0.151	-0.128				
p-value	0.00	0.00	0.00	0.00				

Table 1.7 Effect of EDGAR implementation on trading performance

The table reports coefficients and t-statistics from baseline staggered diff-in-diff and stacked diff-in-diff regressions of the trading performance on the dummy variable Post from Eq. (1.3) and Eq. (1.4). The dependent variable is post-trade sell-buy return differential (*PTSBD*), which is the mean difference between 30-day return following all purchases in each stock in a quarter and 30-day return following all sells in the same stock in the same quarter. The 30-day return is calculated as the dollar amount times excess cumulative returns in 30 days following the trades. *PTSBD* is winsorized at 1 and 99 percentile levels. *Post* is an indicator that equals one if quarter t is on or after the effective date of each implementation phase and zero if quarter t is before that date. The definitions of other explanatory variables are described in Table 1.2. The sample includes ordinary common stocks of U.S. companies listed on major exchanges (NYSE, Amex, and Nasdaq). The sample period is from January 1991 to December 1996. Standard errors are clustered at the stock and year-quarter levels. The t-statistics are reported in parentheses. ***, **, and *indicate statistical significance at the 1%, 5%, and 10% levels, respectively.

	Staggered diff-in-diff analysis: Eq. (1.3)		Stacked diff-in-diff regression: Eq. (1.4)	
	(1)	(2)	(3)	(4)
Post	-31.13**	-26.14*	-69.88*	-73.24**
	(2.42)	(-1.97)	(-2.11)	(-2.24)
Market beta		-14.98		-34.51
		(-0.66)		(-0.98)
Size		25.08		57.19
		(0.91)		(1.45)
MOM		3.691		-19.37
		(0.18)		(-0.49)
B/M		32.54*		12.74
		(1.86)		(0.40)
Price		0.685*		1.581
		(1.80)		(0.99)
Turnover		62.53*		177.1***
		(1.94)		(3.98)
Dividend dummy		-11.97		63.46
		(-0.30)		(0.92)
Nasdaq dummy		-15.58		51.73
		(-0.26)		(0.47)
INST		31.82		7.603
		(0.49)		(0.08)
ACov		3.052		-44.76
		(0.13)		(-0.91)
Bid-ask spread		-1.062		11.14*
		(-0.23)		(1.78)
IVOL		-2.737		-11.20
		(-0.47)		(-1.16)
Controls	No	Yes	No	Yes
Firm FE	Yes	Yes	No	No
Year-quarter FE	Yes	Yes	No	No
Firm FE × Group FE	No	No	Yes	Yes
Year-quarter FE × Group FE	No	No	Yes	Yes
Observations	18,725	18,725	14,069	14,069
Adj R-squared	0.010	0.010	0.001	0.004

Table 1.8 Effect of EDGAR implementation on overconfidence: Excluding the first implementation phase

The table reports the coefficients and *t*-statistics from the staggered and stacked diff-in-diff regressions of the overconfidence on the dummy variable Post with the first implementation phase firms removed from the sample. Panels A and B present the regression results from baseline staggered analysis and stacked analysis, respectively. Dependent variables are four measures of overconfidence. Trading volume (*VO*) is calculated as the natural logarithm of one plus the sum of shares traded (in thousands) in a quarter for stocks that are traded at least once by retail investors in our sample. Dollar Volume (*DVO*) is calculated as the natural logarithm of one plus the sum of dollar shares traded (in thousands) in a quarter. Trading Frequency (*Freq*) is calculated as the natural logarithm of one plus sum of number of trades for each stock in a quarter. Post-trade sell-buy return differential (*PTSBD*) is the mean difference between 30-day return following all purchases in each stock in a quarter and 30-day return following all sells in the same stock in the same quarter. The 30-day return is calculated as the dollar amount times excess cumulative returns in 30 days following the trades. *PTSBD* is winsorized at 1 and 99 percentile levels. The control variables are the same as those in Table 1.2. The sample period is from January 1991 to December 1996. Standard errors are clustered at the stock and year-quarter levels. The *t*-statistics are reported in parentheses. ***, **, and *indicate statistical significance at the 1%, 5%, and 10% levels, respectively.

Panel A: Baseline staggered analysis: Eq. (1.3)

Dependent variable =	Trading volume		Dollar volume		Trading frequency		PTSBD	
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
Post	-0.085***	-0.096***	-0.119***	-0.144***	-0.100***	-0.115***	-33.06**	-27.69**
	(-4.37)	(-5.24)	(-3.66)	(-4.88)	(-4.26)	(-5.19)	(-2.70)	(-2.16)
Controls	No	Yes	No	Yes	No	Yes	No	Yes
Firm FE	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Year-quarter FE	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Observations	33,287	33,287	33,287	33,287	33,287	33,287	17,962	17,962
Adj R-squared	0.742	0.752	0.796	0.810	0.810	0.823	0.008	0.009

Panel B: Stacked diff-in-diff regression: Eq. (1.4)

Dependent variable =	Trading volume		Dollar volume		Trading frequency		PTSBD	
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
Post	-0.144***	-0.139***	-0.191***	-0.206***	-0.131***	-0.136***	-68.90*	-71.96**
	(-4.62)	(-4.84)	(-3.50)	(-4.44)	(-3.82)	(-4.26)	(-2.05)	(-2.14)
Controls	No	Yes	No	Yes	No	Yes	No	Yes
Firm FE	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Year-quarter FE	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Observations	25,817	25,817	25,817	25,817	25,817	25,817	13,742	13,742
Adj R-squared	0.749	0.756	0.801	0.810	0.814	0.823	0.001	0.004

Table 1.9 Pre-trend and falsification tests

The table reports the coefficients and *t*-statistics from the staggered diff-in-diff regressions of the overconfidence on the dummy variable *Post* in the pre-trend assumption and falsification tests with control variables. In the pre-trend assumption (falsification) test, the pseudo-events of each EDGAR implementation phase are set as the date two years before (after) the actual effective date shown in Table 1.1. Accordingly, the dummy variable *Post* equals one if quarter *t* is after the pseudo-event date and zero otherwise. Panel A shows the regression results for the pre-trend test and Panel B shows the results for the falsification test. Dependent variables are four measures of overconfidence. Trading volume (*VO*) is calculated as the natural logarithm of one plus the sum of shares traded (in thousands) in a quarter for stocks that are traded at least once by retail investors in our sample. Dollar Volume (*DVO*) is calculated as the natural logarithm of one plus the sum of dollar shares traded (in thousands) in a quarter. Trading Frequency (*Freq*) is calculated as the natural logarithm of one plus sum of number of trades for each stock in a quarter. Post-trade sell-buy return differential (*PTSBD*) is the mean difference between 30-day return following all purchases in each stock in a quarter and 30-day return following all sells in the same stock in the same quarter. The 30-day return is calculated as the dollar amount times excess cumulative returns in 30 days following the trades. *PTSBD* is winsorized at 1 and 99 percentile levels. The control variables are the same as those in Table 1.2. The sample period is from January 1991 to December 1996. Standard errors are clustered at the stock and year-quarter levels. The *t*-statistics are reported in parentheses. ***, **, and *indicate statistical significance at the 1%, 5%, and 10% levels, respectively.

Panel A: Pre-trends test								
Dependent variable =	Trading volume		Dollar volume		Trading frequency		PTSBD	
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
Post	0.032	0.032	-0.013	0.028	-0.013	0.003	2.457	3.762
	(0.73)	(0.70)	(-0.18)	(0.41)	(-0.34)	(0.07)	(0.12)	(0.20)
Controls	No	Yes	No	Yes	No	Yes	No	Yes
Firm FE	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Year-quarter FE	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Observations	11,250	11,250	11,250	11,250	11,250	11,250	11,249	11,249
Adj R-squared	0.617	0.631	0.616	0.653	0.680	0.707	0.001	0.001

Panel B: Post-trends test								
Dependent variable =	Trading volume		Dollar volume		Trading frequency		PTSBD	
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
Post	0.001	0.036	0.089	0.088	0.050*	0.060**	-28.04	-24.07
	(0.04)	(0.85)	(1.42)	(1.50)	(1.98)	(2.43)	(-0.96)	(-0.80)
Controls	No	Yes	No	Yes	No	Yes	No	Yes
Firm FE	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Year-quarter FE	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Observations	4,857	4,857	4,857	4,857	4,857	4,857	4,856	4,856
Adj R-squared	0.759	0.766	0.722	0.725	0.808	0.811	0.011	0.012

Table 1.10 Effect of EDGAR implementation on the overconfidence: By each implementation phase

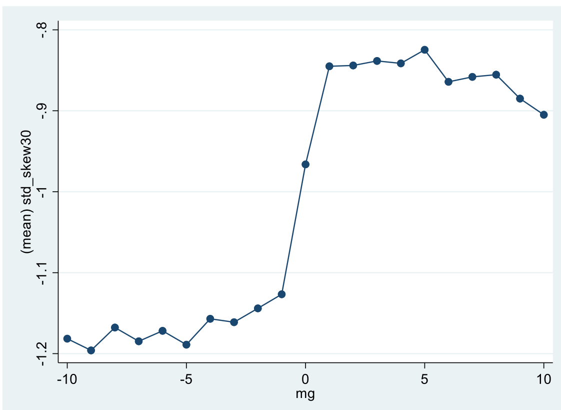
The table reports the coefficients and t-statistics from the diff-in-diff regressions of the overconfidence on the dummy variable *Post*. For each EDGAR implementation phase, we construct the clean 2×2 dataset. The control group contains the firms that have not yet joined the EDGAR platform in the sample. We keep the longest window [-5, 5] for each implementation phase with the first six phases having the control groups. *Post* is the dummy variable that equals one if the year-quarter is on or after the effective implementation year-quarter, and zero otherwise. Dependent variables are four measures of overconfidence. Trading volume (*VO*) is calculated as the natural logarithm of one plus the sum of shares traded (in thousands) in a quarter for stocks that are traded at least once by retail investors in our sample. Dollar Volume (*DVO*) is calculated as the natural logarithm of one plus the sum of dollar shares traded (in thousands) in a quarter. Trading Frequency (*Freq*) is calculated as the natural logarithm of one plus sum of number of trades for each stock in a quarter. Post-trade sell-buy return differential (*PTSBD*) is the mean difference between 30-day return following all purchases in each stock in a quarter and 30-day return following all sells in the same stock in the same quarter. The 30-day return is calculated as the dollar amount times excess cumulative returns in 30 days following the trades. *PTSBD* is winsorized at 1 and 99 percentile levels. The control variables are the same as those in Table 1.2. The sample period is from January 1991 to December 1996. Standard errors are clustered at the stock and year-quarter levels. The t-statistics are reported in parentheses. ***, **, and *indicate statistical significance at the 1%, 5%, and 10% levels, respectively.

Dependent variable =	Trading volume		Dollar volume		Trading frequency		PTSBD	
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
Phase 1	-0.145	-0.211	-0.268	-0.339	-0.107	-0.162	-111.8	-113.5
	(-0.96)	(-1.50)	(-1.03)	(-1.63)	(-0.70)	(-1.04)	(-1.72)	(-1.11)
Phase 2	-0.222**	-0.075	-0.351*	-0.099	-0.205	-0.050	-254.9*	-251.3*
	(-2.54)	(-0.92)	(-1.87)	(-0.55)	(-1.71)	(-0.50)	(-2.17)	(-1.88)
Phase 3	-0.199***	-0.16***	-0.273**	-0.200**	-0.178***	-0.136**	-44.61	-23.49
	(-3.93)	(-3.37)	(-2.92)	(-2.39)	(-3.40)	(-2.87)	(-0.46)	(-0.23)
Phase 4	-0.208***	-0.183***	-0.283**	-0.274***	-0.200***	-0.185***	-3.899	-21.36
	(-3.94)	(-4.01)	(-3.09)	(-3.57)	(-3.64)	(-3.91)	(-0.09)	(-0.52)
Phase 5	-0.102**	-0.116**	-0.140*	-0.188**	-0.092*	-0.117**	-91.62**	-89.21**
	(-2.52)	(-3.12)	(-1.93)	(-3.11)	(-2.03)	(-2.84)	(-2.67)	(-2.41)
Phase 6	-0.083	-0.110**	-0.074	-0.164*	-0.066	-0.112**	-98.49	-102.8
	(-1.66)	(-2.31)	(-0.86)	(-2.07)	(-1.34)	(-2.35)	(-1.72)	(-1.67)
Controls	No	Yes	No	Yes	No	Yes	No	Yes

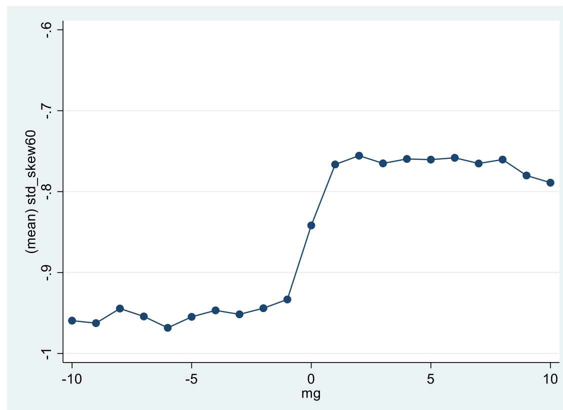
Figure 2.1 Ex-ante skewness in [-10,10] window with bad guidance news disclosure

The charts present the daily average ex-ante skewness within [-10,10] window around the management guidance $t = 0$ when the mangers disclose the shortfall news. Bad news is defined when companies state they are not expected to meet earnings for the period indicated (shortfall), recorded by IBES management guidance database. The ex-ante skewness is estimated following Eq. (2.3) as the variance scaled skewness. The ex-ante skewness using the options with available maturities and interpolated to maturities of 30, 60 and 91 days are shown in Figure1A, 1B and 1C respectively. X-axis represents the day from 10 days prior to the management guidance disclosure day to 10 days post to it, with disclosure day equaling 0. Y-axis represents the average daily ex-ante skewness. The sample includes ordinary common stocks of U.S. nonfinancial companies listed on major exchanges (NYSE, Amex, and Nasdaq) with available option data. The sample period is from January 1996 to December 2022.

A 30 days maturity



B 60 days maturity



C 91 days maturity

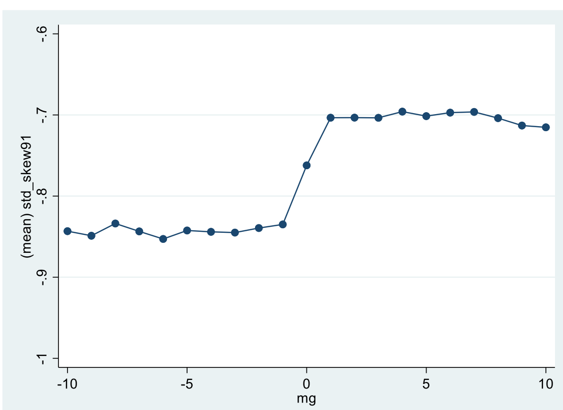
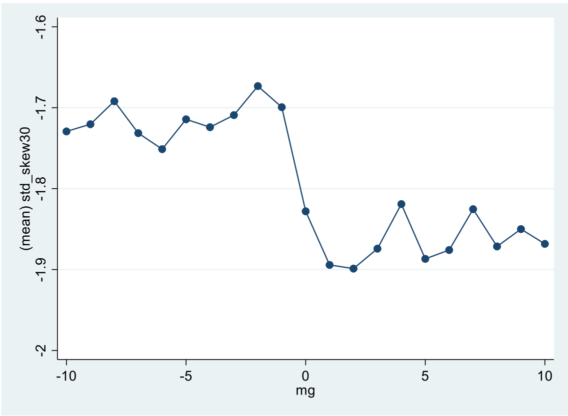


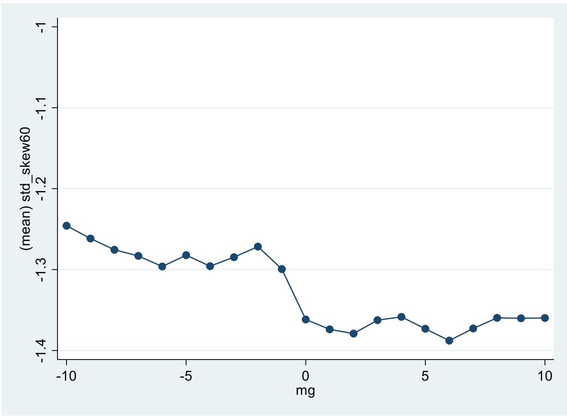
Figure 2.2 Ex-ante skewness in [-10,10] window with good guidance news disclosure

The charts present the daily average ex-ante skewness within [-10,10] window around the management guidance $t = 0$ when the mangers disclose the good news. Good news is defined when companies state they are expected to beat earnings for the period indicated, recorded by IBES management guidance database. The ex-ante skewness is estimated following Eq. (2.3) as the variance scaled skewness. The ex-ante skewness using the options with available maturities and interpolated to maturities of 30, 60 and 91 days are shown in Figure2A, 2B and 2C respectively. X-axis represents the day from 10 days prior to the management guidance disclosure day to 10 days post to it, with disclosure day equaling 0. Y-axis represents the average daily ex-ante skewness. The sample includes ordinary common stocks of U.S. nonfinancial companies listed on major exchanges (NYSE, Amex, and Nasdaq) with available option data. The sample period is from January 1996 to December 2022.

A 30 days maturity



B 60 days maturity



C 91 days maturity

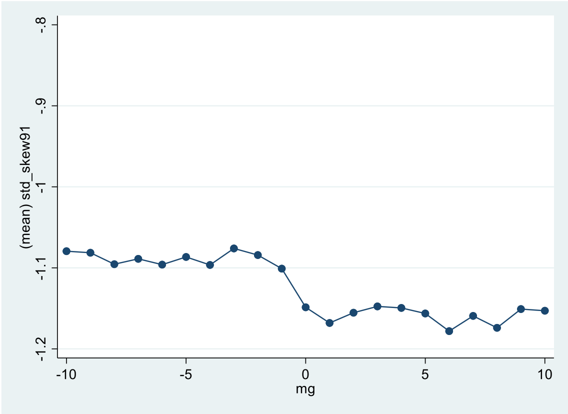


Table 2.1 Summary statistics

This table reports the summary statistics of the variables used in the empirical analysis, including the mean (Mean), standard deviation (Std Dev), and quartiles (P25, P50, and P75). Panel A and Panel B show the main variables in subsamples where the quarterly and annual forecasts are reported. Panel C shows the main variables in full samples and Panel D shows the control variables in full samples. $stdSKEW_{t,t+N}$ is the variance scaled ex-ante skewness calculated following Eq. (2.3), using the options with available maturities and interpolated maturity to N days. $rSKEW_{t,t+N}$ is the realized skewness estimated using the daily stock returns from day t to $t+N$. $DSKEW_{t,t+N}$ is the change of ex-ante skewness, calculated following Eq. (2.4). Good news (*Good*) is defined as dummy variable, which is one if firms' guidance is above the mean of analysts' estimation and zero otherwise. The bad news (*Bad*) is defined as the dummy variable, which is one when firms' guidance is below the mean of analysts' estimation and zero otherwise. Neutral is the dummy variable that equals one if the mean of analysts' estimations falls within the expected range reported by the managers, and zero otherwise. Following are the control variables. *Ret* is the daily stock returns, shown as percentage. *ROA* is calculated as the earnings scaled by total assets. *Leverage* is the ratio of the total liability over the total assets. *B/M* is the natural logarithm of a firm's book equity at the end of the previous fiscal year, divided by the market value of equity at the end of December of the previous year. Firm size (*Size*) is computed as the natural logarithm of the market value of equity, calculated in million dollars. *Turnover* is measured as the daily average number of shares traded in a quarter divided by the number of shares outstanding, shown as a percentage, with the number of shares for Nasdaq stocks adjusted. Market beta (*Beta*) is estimated for each stock using its monthly returns over the prior 60 months. Stock idiosyncratic volatility (*IVOL*) is computed as the standard deviation of the daily residuals in a month from the Fama–French three-factor model. Institutional holdings (*INST*) is measured as the natural logarithm of one plus a stock's institutional ownership scaled by shares outstanding in each quarter. Analyst coverage (*ACov*) is calculated as the natural logarithm of one plus the average number of analysts covering a stock in each quarter. QTR is defined as the dummy variable which is one if the quarterly forecasts are reported and zero if the annual forecasts are reported. Multi_type is defined as the dummy variable which is one when more than one type of guidance measures are reported and zero otherwise. Variables are winsorized at 1% and 99%. The sample includes ordinary common stocks of U.S. nonfinancial companies listed on major exchanges (NYSE, Amex, and Nasdaq). The sample period is from January 1996 to December 2022.

Panel A: Main Variables: Periodicity of Management Guidance--Quarterly

Variable	Mean	SD	P25	P50	P75	N
$DSKEW_{t,t+30}$	0.04	0.87	-0.30	0.02	0.38	8,863
$DSKEW_{t,t+60}$	0.03	0.52	-0.16	0.01	0.21	8,863
$DSKEW_{t,t+91}$	0.03	0.39	-0.10	0.01	0.15	8,863
Bad	0.19	0.39	0.00	0.00	0.00	8,863
Good	0.13	0.33	0.00	0.00	0.00	8,863
Neutral	0.25	0.43	0.00	0.00	1.00	8,863

Panel B: Main Variables: Periodicity of Management Guidance--Annual

Variable	Mean	SD	P25	P50	P75	N
$DSKEW_{t,t+30}$	-0.11	0.98	-0.46	-0.05	0.29	18,060
$DSKEW_{t,t+60}$	-0.04	0.60	-0.23	-0.03	0.17	18,060
$DSKEW_{t,t+91}$	-0.02	0.43	-0.15	-0.01	0.12	18,060
Bad	0.13	0.34	0.00	0.00	0.00	18,060
Good	0.15	0.36	0.00	0.00	0.00	18,060
Neutral	0.47	0.50	0.00	0.00	1.00	18,060

Panel C: Main Variables: Full Sample

Variable	Mean	SD	P25	P50	P75	N
$DSKEW_{t,t+30}$	-0.06	0.95	-0.41	-0.03	0.32	26,923
$DSKEW_{t,t+60}$	-0.02	0.58	-0.21	-0.01	0.18	26,923
$DSKEW_{t,t+91}$	-0.01	0.42	-0.13	0.00	0.13	26,923
Bad	0.15	0.36	0.00	0.00	0.00	26,923

Good	0.15	0.35	0.00	0.00	0.00	26,923
Neutral	0.40	0.49	0.00	0.00	1.00	26,923

Panel D: Control Variables: Full Sample

Variable	Mean	SD	P25	P50	P75	N
Ret %	-0.22	6.13	-1.97	0.03	1.98	26,923
ROA	0.01	0.04	0.00	0.01	0.02	26,923
Leverage	0.21	0.17	0.05	0.19	0.32	26,923
B/M	0.52	0.52	0.24	0.40	0.65	26,923
Size	8.00	1.59	6.86	7.89	9.06	26,923
Turnover %	1.16	2.69	0.56	0.88	1.41	26,923
Beta	1.33	0.78	0.83	1.22	1.72	26,923
IVOL	1.83	1.23	1.03	1.52	2.25	26,923
INST	0.51	0.23	0.47	0.59	0.65	26,923
ACov	2.31	0.62	1.88	2.35	2.77	26,923
QTR	0.33	0.47	0.00	0.00	1.00	26,923
Multi_type	0.18	0.38	0.00	0.00	0.00	26,923

Table 2.2 Ex-ante skewness and realized stock price crash

This table presents the regression results in which the dependent variable is ex-post skewness ($rSKEW_{t,t+N}$) and the independent variable is the $stdSKEW_{t,t+N}$, which is estimated following Eq. (2.3). The regression model is

$$rSKEW_{i,t+1,t+1+N} = \alpha + \beta_1 \times stdSKEW_{i,t,t+N} + \boldsymbol{\beta} \times \boldsymbol{X} + FE_i + FE_t,$$

where $rSKEW_{t,t+N}$ is the realized skewness estimated using the daily stock returns from day t to $t+N$. The ex-ante skewness $stdSKEW_{t,t+N}$ is estimated using the options with available maturities and interpolated to maturities of 30, 60 and 91 days, with N=30,60 and 91. The definitions of other explanatory variables are described in Table 2.1. The sample includes ordinary common stocks of U.S. nonfinancial companies listed on major exchanges (NYSE, Amex, and Nasdaq). The sample period is from January 1996 to December 2022. Standard errors are clustered at the stock and date level. The *t*-statistics are reported in parentheses. ***, **, and *indicate statistical significance at the 1%, 5%, and 10% levels, respectively.

	N=30		N=60		N=91	
	(1)	(2)	(3)	(4)	(5)	(6)
$stdSKEW_{t+N}$	0.017***	0.014***	0.032***	0.026***	0.048***	0.036***
	(15.29)	(12.80)	(11.50)	(9.40)	(9.64)	(7.21)
ROA		-0.569***		-0.78***		-0.817***
		(-8.00)		(-7.04)		(-5.99)
Leverage		-0.024		-0.038		-0.059
		(-1.15)		(-1.17)		(-1.36)
B/M		-0.005		-0.005		-0.005
		(-0.80)		(-0.43)		(-0.34)
Size		-0.114***		-0.177***		-0.220***
		(-23.15)		(-23.16)		(-22.03)
Turnover		-0.004		-0.006*		-0.011**
		(-1.36)		(-1.68)		(-2.25)
Beta		-0.015***		-0.025***		-0.031***
		(-3.15)		(-3.40)		(-3.22)
IVOL		0.006***		0.003		0.005
		(3.16)		(0.98)		(1.47)
INST		-0.006		-0.013		-0.017
		(-0.32)		(-0.42)		(-0.41)
ACov		-0.001		-0.013		-0.021
		(-0.20)		(-1.19)		(-1.43)
Observations	5,581,333	5,581,333	5,581,333	5,581,333	5,581,333	5,581,333
Firm FE	Yes	Yes	Yes	Yes	Yes	Yes
Date FE	Yes	Yes	Yes	Yes	Yes	Yes
Adj R^2	0.04	0.05	0.06	0.04	0.06	0.07

Table 2.3 Ex-ante skewness changes around management guidance

This table presents the regression results in which the dependent variable is the change of ex-ante skewness around the management guidance disclosure date ($DSKEW_{i,t,t+N}$) calculated following Eq. (2.4), with N days maturity of ex-ante skewness. We define good news (*Good*) as dummy variable, which is one if firms' guidance is above the mean of analysts' estimation and zero otherwise. The bad news (*Bad*) is defined as the dummy variable, which is one when firms' guidance is below the mean of analysts' estimation and zero otherwise. Neutral is the dummy variable that equals one if the mean of analysts' estimations falls within the expected range reported by the managers, and zero otherwise. The definitions of other explanatory variables are described in Table 2.1. The coefficients are multiplied by 100. The sample includes ordinary common stocks of U.S. nonfinancial companies listed on major exchanges (NYSE, Amex, and Nasdaq). The sample period is from January 1996 to December 2022. Standard errors are clustered at the stock and date level. The *t*-statistics are reported in parentheses. ***, **, and *indicate statistical significance at the 1%, 5%, and 10% levels, respectively.

	N=30		N=60		N=91	
	(1)	(2)	(3)	(4)	(5)	(6)
Bad	19.93***	12.58***	9.781***	5.550***	7.542***	4.612***
	(7.19)	(4.79)	(5.73)	(3.41)	(6.14)	(3.90)
Good	-10.45***	-5.477**	-4.909***	-2.049	-3.766***	-1.745*
	(-4.05)	(-2.13)	(-3.27)	(-1.38)	(-3.67)	(-1.73)
Neutral	1.213	1.508	-0.520	-0.252	-0.892	-0.710
	(0.64)	(0.81)	(-0.44)	(-0.22)	(-0.98)	(-0.82)
QTR		5.929***		2.947**		1.875**
		(3.11)		(2.42)		(1.98)
Multi_type		3.384*		-0.331		0.172
		(1.91)		(-0.30)		(0.20)
RET		-4.369***		-2.541***		-1.798***
		(-18.66)		(-18.47)		(-17.33)
ROA		50.34***		4.103		7.511
		(2.79)		(0.36)		(0.76)
Leverage		-2.976		-0.139		-2.376
		(-0.41)		(-0.036)		(-0.78)
B/M		-0.230		-0.396		0.474
		(-0.16)		(-0.47)		(0.90)
Size		-2.230		-1.221		-0.110
		(-1.57)		(-1.46)		(-0.18)
Turnover		-1.906**		-1.252**		-0.394
		(-2.00)		(-2.37)		(-0.97)
Beta		0.171		0.413		1.042**
		(0.17)		(0.66)		(2.56)
IVOL		0.654		0.354		0.007
		(0.84)		(0.78)		(0.02)
INST		-6.509		0.574		0.764
		(-0.98)		(0.15)		(0.30)
ACov		0.004		1.075		0.590
		(0.01)		(0.71)		(0.55)
Controls	No	Yes	No	Yes	No	Yes
Observations	26,426	26,426	26,426	26,426	26,426	26,426
Firm FE	Yes	Yes	Yes	Yes	Yes	Yes
Year-quarter FE	Yes	Yes	Yes	Yes	Yes	Yes
Adj R^2	0.064	0.129	0.047	0.106	0.036	0.091

Table 2.4 Ex-ante skewness changes around management guidance: Forecast periodicity

This table presents the regression results in which the dependent variable is the change of ex-ante skewness around the management guidance disclosure date ($DSKEW_{i,t,t+N}$) calculated following Eq. (2.4), with N days maturity of ex-ante skewness. We define good news (*Good*) as dummy variable, which is one if firms' guidance is above the mean of analysts' estimation and zero otherwise. The bad news (*Bad*) is defined as the dummy variable, which is one when firms' guidance is below the mean of analysts' estimation and zero otherwise. Neutral is the dummy variable that equals one if the mean of analysts' estimations falls within the expected range reported by the managers, and zero otherwise. The definitions of other explanatory variables are described in Table 2.1. The coefficients are multiplied by 100. The sample includes ordinary common stocks of U.S. nonfinancial companies listed on major exchanges (NYSE, Amex, and Nasdaq). The sample period is from January 1996 to December 2022. Standard errors are clustered at the stock and date level. The *t*-statistics are reported in parentheses. ***, **, and *indicate statistical significance at the 1%, 5%, and 10% levels, respectively.

	N=30		N=60		N=91	
	(1)	(2)	(3)	(4)	(5)	(6)
Bad×QTR	31.63***	24.55***	19.13***	15.36***	12.27***	9.655***
	(6.86)	(5.50)	(7.11)	(5.91)	(6.10)	(4.93)
Good×QTR	-11.25**	1.631	-3.389	4.501	-2.944	2.743
	(-2.33)	(0.35)	(-1.12)	(1.54)	(-1.29)	(1.22)
Neutral×QTR	-6.524*	-3.284	-1.096	0.908	-1.430	0.095
	(-1.94)	(-1.03)	(-0.54)	(0.46)	(-0.90)	(0.06)
Bad	6.539**	3.456	1.977	-0.003	2.490*	1.103
	(1.99)	(1.07)	(1.04)	(-0.01)	(1.75)	(0.79)
Good	-7.983***	-6.003*	-4.313**	-3.360*	-3.168***	-2.543**
	(-2.61)	(-1.95)	(-2.38)	(-1.86)	(-2.62)	(-2.10)
Neutral	2.088	1.951	-0.667	-0.720	-0.806	-0.89
	(0.94)	(0.91)	(-0.52)	(-0.59)	(-0.84)	(-0.97)
QTR	6.023***	2.291	1.375	-0.791	1.269	-0.311
	(2.88)	(1.17)	(0.95)	(-0.57)	(1.08)	(-0.27)
Multi_type		2.645		-0.770		-0.114
		(1.49)		(-0.70)		(-0.13)
RET		-4.318***		-2.518***		-1.783***
		(-18.52)		(-18.33)		(-17.22)
ROA		51.67***		4.822		7.990
		(2.84)		(0.42)		(0.80)
Leverage		-3.107		-0.163		-2.410
		(-0.43)		(-0.04)		(-0.80)
B/M		-0.131		-0.347		0.506
		(-0.09)		(-0.41)		(0.95)
Size		-2.165		-1.162		-0.075
		(-1.51)		(-1.38)		(-0.12)
Turnover		-1.836*		-1.220**		-0.373
		(-1.92)		(-2.30)		(-0.91)
Beta		0.138		0.443		1.055***
		(0.14)		(0.71)		(2.61)
IVOL		0.793		0.443		0.063
		(1.02)		(0.98)		(0.19)
INST		-6.283		0.738		0.870
		(-0.94)		(0.19)		(0.34)
ACov		0.093		1.160		0.644
		(0.04)		(0.76)		(0.60)

Observations	26,426	26,426	26,426	26,426	26,426	26,426
Firm FE	Yes	Yes	Yes	Yes	Yes	Yes
Year-quarter FE	Yes	Yes	Yes	Yes	Yes	Yes
Adj R^2	0.069	0.131	0.051	0.108	0.039	0.092

Table 2.5 Ex-ante skewness changes around management guidance: Multitype forecasts

This table presents the regression results in which the dependent variable is the change of ex-ante skewness around the management guidance disclosure date ($DSKEW_{i,t,t+N}$) calculated following Eq. (2.4), with N days maturity of ex-ante skewness. We define good news (*Good*) as dummy variable, which is one if firms' guidance is above the mean of analysts' estimation and zero otherwise. The bad news (*Bad*) is defined as the dummy variable, which is one when firms' guidance is below the mean of analysts' estimation and zero otherwise. Neutral is the dummy variable that equals one if the mean of analysts' estimations falls within the expected range reported by the managers, and zero otherwise. The definitions of other explanatory variables are described in Table 2.1. The coefficients are multiplied by 100. The sample includes ordinary common stocks of U.S. nonfinancial companies listed on major exchanges (NYSE, Amex, and Nasdaq). The sample period is from January 1996 to December 2022. Standard errors are clustered at the stock and date level. The *t*-statistics are reported in parentheses. ***, **, and *indicate statistical significance at the 1%, 5%, and 10% levels, respectively.

	N=30		N=60		N=91	
	(1)	(2)	(3)	(4)	(5)	(6)
Bad×Multi_type	26.20***	15.88***	15.66***	9.884***	10.14***	6.135**
	(4.17)	(2.60)	(3.96)	(2.60)	(3.40)	(2.13)
Good×Multi_type	-9.907*	-5.159	-9.142**	-6.137*	-4.431*	-2.198
	(-1.78)	(-0.96)	(-2.56)	(-1.77)	(-1.69)	(-0.86)
Neutral×Multi_ type	-0.148	-2.702	-3.875	-5.201**	-0.310	-1.196
	(-0.04)	(-0.70)	(-1.57)	(-2.18)	(-0.17)	(-0.66)
Bad	15.29***	9.900***	7.080***	3.904**	5.773***	3.578***
	(4.98)	(3.42)	(3.82)	(2.22)	(4.30)	(2.82)
Good	-9.352***	-4.893*	-3.768**	-1.204	-3.278***	-1.489
	(-3.48)	(-1.81)	(-2.38)	(-0.76)	(-3.03)	(-1.39)
Neutral	1.124	2.080	0.266	0.786	-0.827	-0.459
	(0.55)	(1.03)	(0.21)	(0.64)	(-0.89)	(-0.51)
Multi_type	1.122	2.953	0.621	1.572	-0.508	0.108
	(0.36)	(1.02)	(0.32)	(0.84)	(-0.37)	(0.08)
QTR		5.683***		2.790**		1.781*
		(2.99)		(2.31)		(1.88)
RET		-4.348***		-2.526***		-1.789***
		(-18.66)		(-18.45)		(-17.30)
ROA		49.62***		3.668		7.235
		(2.76)		(0.32)		(0.74)
Leverage		-2.702		0.0690		-2.269
		(-0.37)		(0.02)		(-0.75)
B/M		-0.257		-0.405		0.463
		(-0.17)		(-0.48)		(0.87)
Size		-2.156		-1.179		-0.0816
		(-1.51)		(-1.40)		(-0.14)
Turnover		-1.835*		-1.201**		-0.366
		(-1.92)		(-2.27)		(-0.90)
Beta		0.163		0.402		1.038**
		(0.16)		(0.64)		(2.55)
IVOL		0.682		0.374		0.017
		(0.88)		(0.82)		(0.05)
INST		-6.797		0.285		0.649
		(-1.018)		(0.07)		(0.25)
ACov		-0.044		1.029		0.571
		(-0.02)		(0.68)		(0.54)

Observations	26,426	26,426	26,426	26,426	26,426	26,426
Firm FE	Yes	Yes	Yes	Yes	Yes	Yes
Year-quarter FE	Yes	Yes	Yes	Yes	Yes	Yes
Adj R^2	0.065	0.130	0.049	0.107	0.037	0.092

Table 2.6 The rule of analyst coverage

This table presents the regression results in which the dependent variable is the change of ex-ante skewness around the management guidance disclosure date ($DSKEW_{i,t,\mathrm{t+N}}$) calculated following Eq. (2.4), with N days maturity of ex-ante skewness. We define good news (*Good*) as dummy variable, which is one if firms' guidance is above the mean of analysts' estimation and zero otherwise. The bad news (*Bad*) is defined as the dummy variable, which is one when firms' guidance is below the mean of analysts' estimation and zero otherwise. Neutral is the dummy variable that equals one if the mean of analysts' estimations falls within the expected range reported by the managers, and zero otherwise. The definitions of other explanatory variables are described in Table 2.1. The coefficients are multiplied by 100. The sample includes ordinary common stocks of U.S. nonfinancial companies listed on major exchanges (NYSE, Amex, and Nasdaq). The sample period is from January 1996 to December 2022. Standard errors are clustered at the stock and date level. The *t*-statistics are reported in parentheses. ***, **, and *indicate statistical significance at the 1%, 5%, and 10% levels, respectively.

	N=30		N=60		N=91	
	(1)	(2)	(3)	(4)	(5)	(6)
Bad× ACov	-13.03***	-10.75***	-7.412***	-6.040***	-5.220***	-4.271***
	(-3.37)	(-2.91)	(-3.13)	(-2.65)	(-3.03)	(-2.60)
Good× ACov	2.576	-0.614	3.587	1.719	2.014	0.662
	(0.67)	(-0.17)	(1.63)	(0.83)	(1.28)	(0.44)
Neutral× ACov	-0.442	-1.949	1.838	1.060	1.515	0.987
	(-0.15)	(-0.68)	(1.08)	(0.65)	(1.17)	(0.80)
Bad	49.93***	37.25***	26.84***	19.48***	19.54***	14.47***
	(5.05)	(3.95)	(4.34)	(3.28)	(4.28)	(3.32)
Good	-16.54*	-4.274	-13.26**	-6.071	-8.476**	-3.318
	(-1.73)	(-0.47)	(-2.39)	(-1.17)	(-2.08)	(-0.85)
Neutral	1.853	5.669	-5.118	-2.978	-4.682	-3.222
	(0.25)	(0.80)	(-1.12)	(-0.68)	(-1.32)	(-0.96)
ACov	1.662	2.152	1.121	1.263	1.015	0.640
	(0.63)	(0.81)	(0.70)	(0.77)	(0.84)	(0.52)
QTR		6.086***		3.015**		1.908**
		(3.20)		(2.50)		(2.04)
Multi_ type		3.390*		-0.319		0.174
		(1.87)		(-0.28)		(0.20)
RET		-4.203***		-2.447***		-1.728***
		(-24.61)		(-23.89)		(-20.92)
ROA		41.36**		4.537		9.347
		(2.45)		(0.51)		(1.28)
Leverage		-1.455		0.709		-1.437
		(-0.22)		(0.19)		(-0.48)
B/M		-0.229		-0.250		0.401
		(-0.20)		(-0.36)		(0.98)
Size		-1.863		-1.097		-0.030
		(-1.34)		(-1.32)		(-0.05)
Turnover		-1.522*		-1.065**		-0.353
		(-1.83)		(-2.11)		(-0.93)
Beta		0.273		0.480		1.102**
		(0.27)		(0.79)		(2.68)
IVOL		0.492		0.270		-0.016
		(0.66)		(0.62)		(-0.05)
INST		-6.435		0.979		1.176
		(-1.06)		(0.27)		(0.43)

Observations	26,426	26,426	26,426	26,426	26,426	26,426
Firm FE	Yes	Yes	Yes	Yes	Yes	Yes
Year-quarter FE	Yes	Yes	Yes	Yes	Yes	Yes
Adj R^2	0.065	0.130	0.048	0.106	0.037	0.091

Table 2.7 The rule of litigation risk

This table presents the regression results in which the dependent variable is the change of ex-ante skewness around the management guidance disclosure date ($DSKEW_{i,t,t+N}$) calculated following Eq. (2.4), with N days maturity of ex-ante skewness. We define good news (*Good*) as dummy variable, which is one if firms' guidance is above the mean of analysts' estimation and zero otherwise. The bad news (*Bad*) is defined as the dummy variable, which is one when firms' guidance is below the mean of analysts' estimation and zero otherwise. Neutral is the dummy variable that equals one if the mean of analysts' estimations falls within the expected range reported by the managers, and zero otherwise. The definitions of other explanatory variables are described in Table 2.1. The coefficients are multiplied by 100. The sample includes ordinary common stocks of U.S. nonfinancial companies listed on major exchanges (NYSE, Amex, and Nasdaq). The sample period is from January 1996 to December 2022. Standard errors are clustered at the stock and date level. The *t*-statistics are reported in parentheses. ***, **, and *indicate statistical significance at the 1%, 5%, and 10% levels, respectively.

	N=30		N=60		N=91	
	(1)	(2)	(3)	(4)	(5)	(6)
Bad× Litigation	-1.137	-0.975	-0.840	-0.757	-1.080	-1.041
	(-0.59)	(-0.51)	(-0.69)	(-0.65)	(-1.16)	(-1.16)
Good× Litigation	2.917	3.740**	1.289	1.792*	1.717**	2.077**
	(1.57)	(2.10)	(1.20)	(1.68)	(2.10)	(2.54)
Neutral× Litigation	0.919	1.038	1.038	1.120	1.031	1.076*
	(0.73)	(0.85)	(1.28)	(1.43)	(1.61)	(1.73)
Bad	18.50***	11.30***	8.638***	4.488**	6.092***	3.216**
	(5.75)	(3.54)	(4.57)	(2.45)	(4.35)	(2.36)
Good	-6.650**	-0.643	-3.260*	0.249	-1.599	0.893
	(-2.16)	(-0.21)	(-1.96)	(0.15)	(-1.38)	(0.74)
Neutral	2.265	2.619	0.698	1.028	0.340	0.550
	(1.03)	(1.22)	(0.58)	(0.87)	(0.38)	(0.62)
Litigation	-1.597	-1.702	-0.989	-1.147	-0.318	-0.624
	(-1.34)	(-1.44)	(-1.30)	(-1.49)	(-0.56)	(-1.06)
QTR		5.980***		3.007**		1.883**
		(3.16)		(2.50)		(2.03)
Multi_ type		3.437*		-0.321		0.193
		(1.89)		(-0.28)		(0.22)
RET		-4.216***		-2.456***		-1.734***
		(-24.64)		(-23.94)		(-20.96)
ROA		40.43**		3.807		8.966
		(2.36)		(0.43)		(1.22)
Leverage		-2.465		0.202		-1.740
		(-0.37)		(0.05)		(-0.59)
B/M		-0.212		-0.201		0.428
		(-0.18)		(-0.29)		(1.05)
Size		-1.590		-0.905		0.0697
		(-1.14)		(-1.07)		(0.12)
Turnover		-1.670**		-1.144**		-0.439
		(-1.99)		(-2.26)		(-1.14)
Beta		0.533		0.621		1.139***
		(0.52)		(1.01)		(2.70)
IVOL		0.579		0.326		0.0129
		(0.78)		(0.75)		(0.04)
INST		-6.109		1.035		1.221
		(-1.01)		(0.28)		(0.45)
ACov		-0.053		1.025		0.484

		(-0.02)		(0.67)		(0.44)
Observations	26,426	26,426	26,426	26,426	26,426	26,426
Firm FE	Yes	Yes	Yes	Yes	Yes	Yes
Year-quarter FE	Yes	Yes	Yes	Yes	Yes	Yes
Adj R^2	0.064	0.128	0.047	0.105	0.036	0.091

Table 2.8 Predicting the realized short-term skewness

this table presents the regression results in which the independent variable of interest is the change of ex-ante skewness around the management guidance disclosure date ($DSKEW_{i,t,t+N}$) calculated following Eq. (2.4), with N days maturity of ex-ante skewness. The regression model is

$$rSKEW_{i,t+3,t+33} = \alpha + \beta_1 \times DSKEW_{i,t,t+N} + \beta_2 \times stdSKEW_{i,t-3,t-3+N} + \beta_3 \times rSKEW_{i,t-31,t-1} + \boldsymbol{\beta} \times \boldsymbol{X} + FE_i + FE_t$$

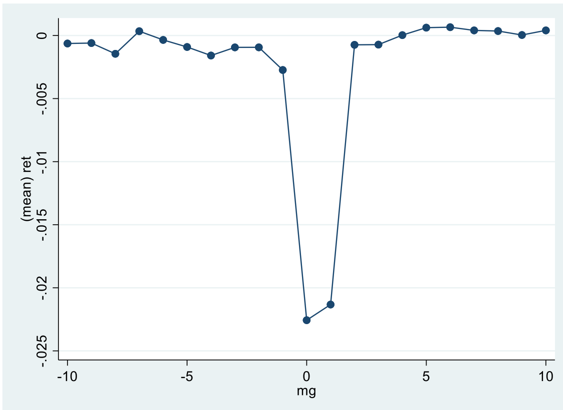
The definitions of other explanatory variables are described in Table 2.1. The sample includes ordinary common stocks of U.S. nonfinancial companies listed on major exchanges (NYSE, Amex, and Nasdaq). The sample period is from January 1996 to December 2022. Standard errors are clustered at the stock and date level. The *t*-statistics are reported in parentheses. ***, **, and *indicate statistical significance at the 1%, 5%, and 10% levels, respectively.

	N=30	
	(1)	(2)
$DSKEW_{t+N}$	0.01*	0.011*
	(1.75)	(1.96)
$stdSKEW$	0.016***	0.014***
	(3.65)	(3.45)
$rSKEW$	-0.008	-0.007
	(-1.24)	(-1.17)
Ret		0.001
		(0.90)
QTR		-0.017
		(-0.99)
Multi_type		-0.001
		(-0.03)
ROA		-0.232
		(-1.30)
Leverage		-0.076
		(-1.09)
B/M		-0.006
		(-0.28)
Size		-0.095***
		(-6.64)
Turnover		0.001
		(0.10)
Beta		0.005
		(0.40)
IVOL		0.003
		(0.38)
INST		-0.081
		(-1.61)
ACov		0.004
		(0.18)
Observations	25,908	25,908
Firm FE	Yes	Yes
Year-quarter	Yes	Yes
Adj R^2	0.051	0.054

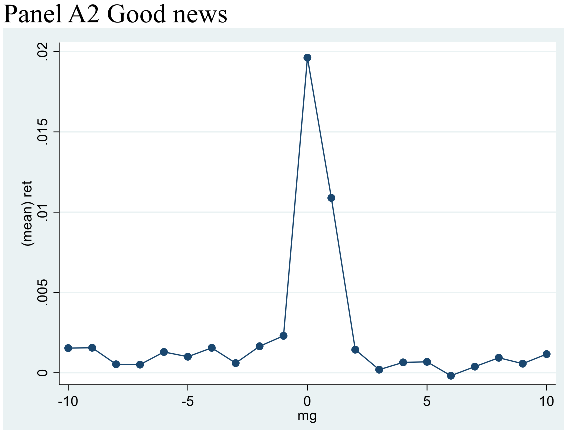
Appendix

Appendix A
The daily stock returns within [-10,10] window around earnings forecasts.

Panel A1 Bad News



Panel A2 Good news



Appendix B. Ex-ante skewness changes around management guidance: firm size

This table presents the regression results in which the dependent variable is the change of ex-ante skewness around the management guidance disclosure date ($DSKEW_{i,t,t+N}$) calculated following Eq. (2.4), with N days maturity of ex-ante skewness. We define good news (*Good*) as dummy variable, which is one if firms' guidance is above the mean of analysts' estimation and zero otherwise. The bad news (*Bad*) is defined as the dummy variable, which is one when firms' guidance is below the mean of analysts' estimation and zero otherwise. Neutral is the dummy variable that equals one if the mean of analysts' estimations falls within the expected range reported by the managers, and zero otherwise. Firm size (*Size*) is computed as the natural logarithm of the market value of equity, calculated in million dollars. The definitions of other explanatory variables are described in Table 2.1. The coefficients are multiplied by 100. The sample includes ordinary common stocks of U.S. nonfinancial companies listed on major exchanges (NYSE, Amex, and Nasdaq). The sample period is from January 1996 to December 2022. Standard errors are clustered at the stock and date level. The *t*-statistics are reported in parentheses. ***, **, and *indicate statistical significance at the 1%, 5%, and 10% levels, respectively.

	N=30		N=60		N=91	
	(1)	(2)	(3)	(4)	(5)	(6)
Bad× Size	-5.106***	-3.567***	-3.264***	-2.363***	-2.045***	-1.382**
	(-3.82)	(-2.70)	(-4.15)	(-3.10)	(-3.47)	(-2.42)
Good×Size	1.352	-1.077	0.670	-0.745	0.817	-0.161
	(0.90)	(-0.75)	(0.81)	(-0.95)	(1.46)	(-0.30)
Neutral× Size	0.675	-0.446	1.642***	1.053*	1.122**	0.743*
	(0.63)	(-0.42)	(2.79)	(1.87)	(2.58)	(1.80)
Bad	59.80***	40.49***	35.22***	24.00***	23.48***	15.41***
	(5.42)	(3.70)	(5.30)	(3.73)	(4.62)	(3.12)
Good	-21.29*	3.013	-10.27	3.869	-10.32**	-0.474
	(-1.73)	(0.25)	(-1.48)	(0.59)	(-2.14)	(-0.10)
Neutral	-4.693	4.738	-14.22***	-9.143*	-10.25***	-6.935*
	(-0.52)	(0.54)	(-2.72)	(-1.84)	(-2.60)	(-1.86)
Size	1.405	-0.913	0.368	-0.900	0.845	-0.01
	(1.04)	(-0.62)	(0.44)	(-1.03)	(1.42)	(-0.01)
QTR		6.070***		2.992**		1.908**
		(3.20)		(2.49)		(2.05)
Multi_type		3.274*		-0.449		0.102
		(1.81)		(-0.39)		(0.12)
RET		-4.201***		-2.444***		-1.726***
		(-24.46)		(-23.78)		(-20.84)
ROA		41.33**		4.344		9.253
		(2.43)		(0.49)		(1.26)
Leverage		-1.724		0.512		-1.604
		(-0.26)		(0.13)		(-0.54)
B/M		-0.243		-0.205		0.424
		(-0.21)		(-0.29)		(1.04)
Turnover		-1.675**		-1.118**		-0.390
		(-2.00)		(-2.23)		(-1.02)
Beta		0.288		0.487		1.104***
		(0.28)		(0.81)		(2.69)
IVOL		0.508		0.220		-0.0440
		(0.69)		(0.51)		(-0.14)
INST		-6.323		1.320		1.398
		(-1.03)		(0.36)		(0.51)
ACov		-0.145		0.892		0.397
		(-0.06)		(0.58)		(0.36)

Observations	26,426	26,426	26,426	26,426	26,426	26,426
Firm FE	Yes	Yes	Yes	Yes	Yes	Yes
Year-quarter FE	Yes	Yes	Yes	Yes	Yes	Yes
Adj R^2	0.065	0.130	0.049	0.106	0.037	0.091

Appendix C.
Predicting the realized short-term skewness

This table presents the regression results in which the independent variable of interest is the change of ex-ante skewness around the management guidance disclosure date ($DSKEW_{i,t,t+N}$) calculated following Eq. (2.4), with N days maturity of ex-ante skewness. The regression model is

$$rSKEW_{i,t+3,t+33} = \alpha + \beta_1 \times DSKEW_{i,t,t+N} + \beta_2 \times stdSKEW_{i,t-3,t-3+N} + \beta_3 \times rSKEW_{i,t-31,t-1} + \boldsymbol{\beta} \times \boldsymbol{X} + FE_i + FE_t$$

The definitions of other explanatory variables are described in Table 2.1. The sample includes ordinary common stocks of U.S. nonfinancial companies listed on major exchanges (NYSE, Amex, and Nasdaq). The sample period is from January 1996 to December 2022. Standard errors are clustered at the stock and date level. The *t*-statistics are reported in parentheses. ***, **, and *indicate statistical significance at the 1%, 5%, and 10% levels, respectively.

	N=60		N=91	
	(1)	(2)	(3)	(4)
$DSKEW_{t+N}$	0.012	0.007	0.020	0.011
	(1.03)	(0.57)	(0.93)	(0.50)
$stdSKEW$	0.020**	0.017**	0.048***	0.036***
	(2.48)	(2.15)	(3.42)	(2.60)
$rSKEW$	-0.011	-0.010	-0.028***	-0.025***
	(-1.43)	(-1.25)	(-3.54)	(-3.14)
Ret		-0.002		-0.003*
		(-1.59)		(-1.70)
QTR		0.012		0.045*
		(0.63)		(1.96)
Multi_type		0.005		-0.015
		(0.23)		(-0.64)
ROA		-0.673**		-0.679**
		(-2.53)		(-2.22)
Leverage		-0.133		-0.039
		(-1.52)		(-0.39)
B/M		-0.006		0.008
		(-0.04)		(0.41)
Size		-0.146***		-0.218***
		(-8.04)		(-9.77)
Turnover		-0.013		-0.024*
		(-0.99)		(-1.71)
Beta		-0.011		-0.011
		(-0.72)		(-0.63)
IVOL		0.012		0.013
		(1.18)		(1.21)
INST		0.062		0.035
		(0.86)		(0.41)
ACov		-0.011		0.020
		(-0.40)		(0.58)
Observations	25,908	25,908	26,031	26,031
Firm FE	Yes	Yes	Yes	Yes
Year-quarter	Yes	Yes	Yes	Yes
Adj R^2	0.055	0.061	0.060	0.069