Journal of Empirical Finance 87 (2026) 101709

Contents lists available at ScienceDirect

Journal of Empirical Finance

journal homepage: www.elsevier.com/locate/jempfin

What drives retail investors' overconfidence? The role of information acquisition costs[☆]

Gang Li [a], Shuqi Wang [b], K.C. John Wei [a,*]

[a] Hong Kong Polytechnic University, Hong Kong, People's Republic of China
[b] University of Stavanger, Stavanger, Norway

ARTICLE INFO

JEL classification:
C13
C53
G12
G14

Keywords:
EDGAR implementation
Overconfidence
Information acquisition costs
Behavioral biases

ABSTRACT

We examine whether information acquisition costs affect the overconfidence of retail investors. Using the implementation of the Electronic Data Gathering, Analysis, and Retrieval (EDGAR) system as an exogenous event, we find that overconfidence, measured by investors' post-trade performance, is significantly reduced after companies join the EDGAR platform, and the reduction is greater for companies with higher information uncertainty. The decrease in overconfidence after the implementation of EDGAR is associated with lower levels of investor disagreement and stock mispricing.

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 et al., 1977). Models of financial markets with overconfident traders imply high trading volume, high volatility, and low asset price informativeness, and help explain market

☆ We gratefully acknowledge Kewei Hou (the editor) and the anonymous referee for constructive and insightful comments and suggestions that have significantly improved the paper. We also thank Terrance Odean and Li An for generously providing us with the trading data. We also thank Itzhak Ben-David, J.C. Lin, Jeff Ng, Alberto Rossi, and conference participants at the 2023 Conference on Pacific Basin Finance, Economics, Accounting, and Management (PBFEAM) and the 2023 Asian Finance Association (Asian FA) conference for helpful comments and suggestions. John Wei acknowledges financial support from the Research Grants Council of the Hong Kong Special Administrative Region, China (GRF15500423). All remaining errors are our own.

* Corresponding author at: School of Accounting and Finance, Hong Kong Polytechnic University, Hung Hom, Kowloon, Hong Kong, China.
E-mail addresses: Gang.G.Li@Polyu.Edu.Hk (G. Li), shuqi.wang@uis.no (S. Wang), Johnwei@ust.hk, john.wei@polyu.edu.hk (K.C.J. Wei).

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

https://doi.org/10.1016/j.jempfin.2026.101709
Received 24 April 2025; Received in revised form 1 March 2026; Accepted 13 March 2026
Available online 17 March 2026
0927-5398/

anomalies, which are strongly supported by empirical and experimental evidence.[2] Studies on what causes retail investors to become overconfident are limited.

In this paper, we examine how increased access to financial information affects the overconfidence of retail investors. On the one hand, the self-attribution bias theory suggests that more financial information may increase overconfidence among retail investors. According to Daniel et al. (1998), investors with biased self-attribution tend to internalize evidence that confirms their preexisting beliefs, further boosting their confidence while dismissing or underweighting contradictory evidence. Consistent with this theory, Chou et al. (2021) show that institutional investors become increasingly overconfident when earnings announcements confirm their previous predictions about stock performance. The broader literature on biased self-attribution documents similar patterns. Statman et al. (2006) link self-attribution to higher trading volume and return volatility. Chui et al. (2010) find that cultural factors amplify overconfidence and momentum profits in more individualistic societies. Similarly, Adam et al. (2015) show that managerial overconfidence increases speculative activity after successful outcomes. These findings suggest that greater access to financial information reinforces investor overconfidence through self-attribution bias.

However, greater information availability may reduce overconfidence by providing investors with more comprehensive data to consider before making investment decisions. Behavioral biases tend to intensify when individuals face complex problems with limited feedback or ambiguous signals (Kahneman, 2003; Kahneman and Tversky, 1973). Theoretical models emphasize that biased judgment is more likely when valuation information is scarce or noisy (Hirshleifer, 2001). Empirical evidence also points in this direction. Kumar (2009) shows that higher uncertainty in both stock and market environments amplifies behavioral biases. In this context, improved access to firm-level fundamentals may help retail investors make better-informed decisions, thereby reducing their reliance on subjective or overconfident beliefs. From this perspective, a more transparent and accessible information environment could serve to mitigate, rather than amplify, investor overconfidence.

We examine whether increased access to financial information strengthens or weakens retail investor overconfidence. 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 to examine its impact on retail investor overconfidence. Previous studies show that even for public information, such as financial statements, acquisition and analysis costs are not negligible and can influence investor behavior and market outcomes.[3] By reducing these costs, EDGAR effectively increases the access to information for retail investors.

We choose the implementation of EDGAR as an exogenous shock to information acquisition costs for three reasons. First, the implementation was mandatory and was carried out 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 about endogeneity. Second, EDGAR substantially reduced the information acquisition costs of accessing corporate disclosures by replacing inefficient paper-based filings with instant electronic access, fundamentally changing how investors obtain firm information and improving transparency.[4] Third, both anecdotal and empirical evidence indicates that retail investors actively use EDGAR. Gao and Huang (2020) document that approximately 31 % of all download requests during the implementation period came from retail investors. In summary, the SEC's randomized implementation, the significant drop in access costs, and the extensive usage by retail investors support the use of EDGAR as a valid exogenous shock to the investor information environment.

We examine the impact of reduced information acquisition costs on the overconfidence of retail investors using a database that contains over 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. We measure the overconfidence of retail investors using the post-trade performance following Kumar (2009). Overconfident investors often exhibit systematic errors in stock valuation as they tend to overestimate the accuracy of their private information, resulting in poorer investment outcomes. Consequently, the stocks they sell tend to outperform the stocks they buy following their trades. We capture this behavior using the 30-day post-trade sell-buy return differential (*PTSBD*), where higher values indicate greater overconfidence.[5]

EDGAR implementation was conducted in ten phases between 1993 and 1996. This staggered implementation allows us to adopt a stacked difference-in-differences analysis ("diff-in-diff" hereafter), following Cengiz et al. (2019), to examine the effect of EDGAR implementation on investor overconfidence. This approach improves traditional staggered diff-in-diff models by constructing clean,

[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 (1998b) 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 that overconfident individual traders continue to trade despite their poor performance. Barber and Odean (2002) later show supporting evidence that overconfidence can explain the increase in trading and the reduction in the performance of online investors. Experimental evidence suggests that investors are more likely to perform worse in trading when they overestimate the precision of their signals (Biais, Hilton, Mazurier, and Pouget, 2005). Glaser and Weber (2007) show that overconfident investors trade more using the survey data.

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

[4] Before EDGAR, firms were required to submit paper copies of financial disclosures to the SEC, making information access inefficient and prone to loss. EDGAR enabled companies to electronically upload financial statements, providing investors with convenient and free online access.

[5] Another popular proxy for overconfidence is trading volume. However, trading volume is noisier as it also reflects investor disagreement (Daniel et al., 1998; Odean, 1998b; Gervais and Odean, 2001; Scheinkman and Xiong, 2003; Grinblatt and Han, 2005). We do not use trading volume to measure overconfidence here. However, we do find that trading volume declines after EDGAR implementation, consistent with Chang, Hsiao, Ljungqvist, and Tseng (2022).

event-specific 2×2 datasets that isolate the treatment effect in each implementation phase. Specifically, for each EDGAR phase, we define an event window from four quarters before to four quarters after implementation ([−4, 4]) around the shock quarter, using only firms that remain untreated at the end of the window as the control group. For example, for the implementation of phases 2–4, which is effective in 1994Q1, valid control firms are drawn from later phases (e.g., phases 7–10), whose implementation dates begin after 1995Q1. These phase-level datasets are then stacked to form the sample that we examine. This stacked diff-in-diff approach mitigates potential biases from treatment effect heterogeneity and contamination across treated and control groups, issues that often affect standard staggered designs.

Our empirical results show that the reduction in information acquisition costs resulting from the implementation of EDGAR significantly decreases the overconfidence of retail investors. On average, *PTSBD* decreases by 1.233 %, which is approximately 4.57 times its mean and 14.56 % of its standard deviation, indicating a significant improvement in trading performance and a reduction in the overconfidence of retail investors after the implementation of EDGAR. We also test parallel pre-treatment trends between treated and control firms and find no statistically significant differences in *PTSBD* during the pre-treatment quarters.

To further show that the reduction in investor overconfidence is driven by lower information acquisition costs and greater access to firm-level data, we perform a subsample analysis using proxies for information uncertainty. We expect the impact of EDGAR to be more pronounced for stocks with high information uncertainty and those that are more difficult to value. We use four firm characteristics to proxy for information uncertainty: firm age, book-to-market equity, idiosyncratic volatility, and Nasdaq listing status. Younger firms, with shorter public histories, offer less established reporting records and are associated with lower information quality (Barry and Brown, 1985; Leary and Roberts, 2010; Maskara and Mullineaux, 2011). Firms with low book-to-market (growth) firms generally disclose less and are opaquer than value firms (Gao and Liang, 2013; Zhang, 2006). Firms with higher idiosyncratic volatility are harder to value (Kumar, 2009). Nasdaq-listed companies, typically in high-technology sectors, present greater information complexity, requiring more specialized knowledge for average retail investors (Chen et al., 2025).

We divide the sample into two sub-samples based on these firm characteristics observed prior to the EDGAR event window and estimate the effects of the EDGAR implementation separately for each sub-sample. Across all sub-samples, we find that the decline in retail investor overconfidence is consistently more pronounced for the subsample of firms with higher information uncertainty. These results support the interpretation that EDGAR implementation reduces the overconfidence of retail investors by making firm fundamental information more accessible.

We also examine the implications of overconfidence in connection with recent studies on the implementation of EDGAR. Chang et al. (2022) show that investor disagreement declines following the adoption of EDGAR. We implement a two-stage least squares (2SLS) approach to examine whether the decline in disagreement is attributed to the decrease in overconfidence. Specifically, we regress investor disagreement, proxied by retail trading volume, on the *PTSBD* instrumented by the EDGAR post dummy, which captures the EDGAR-induced variation in overconfidence. We find that the decline in disagreement following the implementation of EDGAR is associated with the decline in the overconfidence of retail investors, consistent with the notion that overconfident investors overweight their own signals, leading to the dispersion of beliefs (Einhorn, 1980; Griffin and Tversky, 1992; Xiong, 2013). Overconfidence can also distort asset prices when investors overreact to private signals and arbitrage is limited (Daniel et al., 1998; Shleifer and Vishny, 1997). Hirshleifer and Ma (2024) demonstrate that EDGAR implementation reduces stock mispricing. We follow their approach and examine whether EDGAR specifically reduces overconfidence-driven mispricing. We first show that the mispricing measure of Stambaugh and Yuan (2017) predicts future stock returns more strongly when overconfidence is higher during our sample period. We then include a triple interaction among overconfidence, the mispricing measure, and the EDGAR post dummy. The coefficient is negative albeit statistically nonsignificant, and the magnitude of the coefficient is economically meaningful, indicating that the implementation reduces the pre-EDGAR overconfidence-driven mispricing.

We perform several robustness checks to validate our findings. First, we perform a falsification test, using the implementation dates of the second, third, and fourth EDGAR phases in 1993 as the treatment dates when filings in EDGAR were not available online, i.e., the filings were not freely accessible to retail investors. As expected, there is no significant change in overconfidence around these placebo treatment dates. Second, to ensure that our results are not driven by any single EDGAR phase, we estimate separate diff-in-diff regressions for each phase. The decline in overconfidence appears in nearly all phases, although magnitude and significance vary. Together, these results confirm that the effect of EDGAR implementation is robust and not phase-specific.

Our paper contributes to the literature on the determinants of overconfidence by focusing on the external information environment. Prior work links investor overconfidence to cognitive biases such as self-attribution (Daniel et al., 1998), overestimation of private signal precision and their own ability (Kyle and Wang, 1997), and biological, cultural, and social influences, including gender and individualistic norms (Barber and Odean, 2001). In contrast to these internal or demographic factors, our findings highlight a different mechanism, the structure of the information environment. We show that when the cost of acquiring financial information is lower, retail investors exhibit significantly less overconfidence.

Our study also contributes to the literature on the effects of EDGAR implementation by highlighting its impact on overconfidence. Previous studies indicate that the implementation of EDGAR improves information flow and market efficiency. For example, Kim et al. (2021) find a causal relation between information costs and accounting-related stock anomalies. Goldstein et al. (2023) find that EDGAR leads to lower capital costs, increased equity fundraising and investment, and reduced managerial reliance on stock prices. Chang et al. (2023) show that after the implementation of EDGAR, analysts significantly reduce coverage and issue less optimistic, more accurate forecasts. Our findings align with those of Gao and Huang (2020), who show that net stock buying by individual investors after earnings announcements is linked to better subsequent cumulative abnormal returns after the EDGAR implementation. We differ by directly measuring investors' actual trading outcomes using the return differential between stocks that investors sell and buy. Our results also complement Chang et al. (2022), who show that EDGAR reduces investor disagreement. We extend their findings

Table 1
The timetable for the EDGAR system implementation.

Implementation phase	Implementation date	Shock date	Effective year-quarter
1	April 26, 1993	January 17, 1994	1994Q1
2	July 19, 1993	January 17, 1994	1994Q1
3	October 4, 1993	January 17, 1994	1994Q1
4	December 6, 1993	January 17, 1994	1994Q1
5	January 30, 1995	January 30, 1995	1995Q1
6	March 6, 1995	March 6, 1995	1995Q1
7	May 1, 1995	May 1, 1995	1995Q2
8	August 7, 1995	August 7, 1995	1995Q3
9	November 6, 1995	November 6, 1995	1995Q4
10	May 6, 1996	May 6, 1996	1996Q2

This table shows the EDGAR implementation timelines. The SEC assigns U.S. public firms into ten groups to join the EDGAR platform mandatorily from 1993 to 1996. We obtain the implementation date from SEC Release 33–6977 and SEC Release 33–7122. The shock date is set differently for the first four groups as online access to EDGAR becomes available on January 17, 1994. The effective year-quarter corresponds to the shock date.

by identifying reduced overconfidence as a key channel. We build on Hirshleifer and Ma (2024) by showing that EDGAR reduces not only mispricing, but specifically the overconfidence-driven component of mispricing. Finally, we make a methodological contribution by adopting the stacked diff-in-diff approach with matched controls (Cengiz et al. 2019), which is well-suited to the staggered implementation of EDGAR experiments.

2. The EDGAR system and retail investors

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

2.1. The EDGAR system and information acquisition for 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 submit financial filings electronically, allowing market participants to access and download disclosures for free. The system significantly reduced information acquisition costs, particularly for retail investors.

Before EDGAR, the disclosure and 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 in person. After being reviewed by the SEC, those paper copies were stored in public reference rooms in Washington D.C., New York, and Chicago, where access was limited by location, availability, and checkout restrictions. As a result, it was time consuming or even impossible for investors, especially retail investors, to acquire the information in the financial statements promptly. Although retail investors can request paper copies of a company's financial statements by mail directly from the company, delivery delays and the difficulty of comparing multiple firms' statements create additional barriers.

EDGAR addressed these inefficiencies by digitizing disclosures and providing same-day public access to filings. 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. 14,628 (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." The SEC's annual report confirms that 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 EDGAR to acquire information when the system was launched. 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.

2.2. The staggered EDGAR implementation

The EDGAR implementation was conducted in ten phases between 1993 and 1996, with the SEC randomly assigning all U.S. public firms to ten different phases. Each group was required to begin electronic filing during its assigned implementation period. Following an initial pilot phase involving voluntary submissions, mandatory implementation was carried out over three years.[6] According to 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 final phase (Group CF-10) were required to file in May 1996.

Table 1 reports the complete EDGAR phase-in schedule, including the implementation and shock dates used in our study. Retail

[6] Before the SEC began the mandatory EDGAR implementation for all U.S. public firms, it invited companies to voluntarily submit filings online during a pilot phase in the 1980s. Firms in the first implementation phase included those who had opted into this voluntary program.

investors gained free access to EDGAR filings starting January 17, 1994, through a public dissemination project led by the Internet Multicasting Service and hosted by New York University. This date marks the point at which EDGAR filings became widely accessible online. We follow the literature in using this date as the shock point for firms in the first four phases. For phases five through ten, the shock date coincides with each group's official implementation date. We exclude the first phase from the analysis because its implementation date, April 26, 1993, occurred six months before the shock date. Additionally, phase one includes firms that joined EDGAR voluntarily during its pilot phase in the 1980s, raising concerns about the randomized assignment assumption needed for our diff-in-diff analysis.

3. Data and measures of the variables

This section discusses our sample selection and measures of the overconfidence and control variables.

3.1. Sample selection

We use the trading data of retail investors at a major U.S. discount brokerage, first introduced by Odean (1998a). The raw dataset spans January 1991 to December 1996 and covers every trade executed through the brokerage, including an account identifier, security identifier, trade date, buy or sell indicator, quantity, price, and CUSIP. This dataset has been widely used in the literature to examine retail investors' trading behaviors, including overconfidence and the disposition effect.[7] We restrict the sample to common equities of U.S. corporations (share classes 10 and 11) and listed on the New York Stock Exchange, American Stock Exchange, or Nasdaq National Market System (exchange codes 1, 2, or 3) and exclude financial services firms. Our empirical analysis examines the period from January 1993 to January 1996 around the EDGAR shock dates and required by our stacked diff-in-diff design.

Following Ben-David and Hirshleifer (2012), we reconstruct investor positions using trade-level data. If an investor holds a negative cumulative position in a given stock, we remove all trades involving that stock by the investor, as negative positions may indicate either pre-sample holdings or short selling, introducing ambiguity.[8] The EDGAR implementation may have attracted more sophisticated investors who began trading after gaining access to firm disclosures. To mitigate this selection bias, we restrict our sample to the investors who initiated trading activity prior to the EDGAR implementation, ensuring a comparable investor base throughout the analysis period.

We obtain financial statement data from Compustat, stock market data from the Center for Research in Security Prices (CRSP), and analyst coverage data from the Institutional Brokers' Estimate System (I/B/E/S).

3.2. Measuring overconfidence

Overconfident investors tend to overestimate the precision of their information or their ability to interpret it, often leading to systematic trading errors (Daniel et al., 1998). An observable outcome is that the stocks they sell systematically outperform the ones they buy. Following this idea, Kumar (2009) measures investor overconfidence as the return difference between the stocks investors sell and the stocks they buy. The study finds that this performance difference tends to be larger when there is higher valuation uncertainty in stocks, indicating a greater degree of overconfidence.

We follow Kumar (2009) and use the 30-day post-trade performance differential at the stock-quarter level as the proxy for the overconfidence of retail investors. The 30-day post-trade sell-buy return differential for stock i in each quarter t, $PTSBD_{i,t}$, is defined as the average difference between the 30-day returns following all sell trades and those following all buy trades of stock i during quarter t.[9] Higher values of $PTSBD_{i,t}$ indicate greater overconfidence, while a decline in this measure reflects a reduction in overconfident trading behavior.

3.3. Control variables

We use similar control variables as those in Kumar (2009) that have been shown to influence investor overconfidence. We begin by controlling exposure to systematic risk. The controls 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 calculated as the natural logarithm of the firm's market value of equity in million dollars at the end of the previous quarter. *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 the last December. *Mom* is the natural logarithm of one plus the cumulative return of a stock in the last 12 months, excluding the most recent month. Kumar (2009) shows that overconfidence is stronger for stocks which are speculative and difficult to assess, such as Nasdaq firms, non-dividend payers, and low-price stocks. We account for these firm characteristics by adding a Nasdaq

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

[8] The trading data of retail investors have been available since January 1991 and our main analysis uses the sample from January1993 onward. The extra two-year data prior to the period of our analysis facilitate the identification of short selling.

[9] For each trade, we compute the 30-day buy-and-hold return from the day after the trade. At the stock–month level, we take the difference between the dollar volume-weighted average 30-day returns on sell trades and on buy trades. The quarterly *PTSBD* for stock i in quarter t is the average of these monthly differentials over the months in quarter t.

Table 2
Summary statistics.

	Mean	Std Dev	P25	P50	P75	N
PTSBD	−0.27	8.47	−3.94	−0.16	3.60	9445
Beta	1.42	0.96	0.83	1.32	1.89	9445
Size	5.16	1.29	4.27	5.13	6.00	9445
Mom	0.07	0.47	−0.20	0.07	0.35	9445
B/M	−0.89	0.86	−1.41	−0.85	−0.30	9445
Price	17.00	12.86	7.50	14.00	23.00	9445
Turnover	0.43	0.41	0.17	0.29	0.54	9445
Dividend dummy	0.33	0.47	0.00	0.00	1.00	9445
Nasdaq dummy	0.60	0.49	0.00	1.00	1.00	9445
Inst	0.08	0.15	0.00	0.00	0.04	9445
ACov	1.21	0.75	0.69	1.20	1.79	9445
Bid-ask spread	3.34	2.59	1.66	2.64	4.19	9445
IVol	2.97	1.55	1.96	2.69	3.60	9445
TV	1.53	1.04	0.69	1.38	2.20	9445
Age	11.39	12.11	3.23	7.10	15.01	9445

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). Post-trade sell-buy return differential (*PTSBD*) is the quarterly mean of the difference between the 30-day buy-and-hold returns following all sell trades and those following all buy trades for a given stock. *Beta* is estimated using 60-month rolling-window regressions of excess returns on the market returns. Firm size (*Size*) is the natural logarithm of the market value of equity 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 the natural logarithm of book-to-market equity. *Price* is the quarter-end stock price. *Turnover* is 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* is one if the stock is listed on the Nasdaq exchange. Institutional ownership (*Inst*) is the natural logarithm of one plus institutional ownership scaled by shares outstanding for each stock at quarter end. Analyst coverage (*ACov*) is 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 a quarter. Idiosyncratic volatility (*IVol*) is the standard deviation of the daily stock return residuals from the Fama-French three-factor model. Trading volume (*TV*) is the natural logarithm of one plus the sum of shares traded (in thousands) in a quarter. Firm age (*Age*) is the number of years since the firm was first covered by the CRSP. All variables are at the quarterly frequency.

indicator, a dividend indicator, and the level of stock prices as control variables. The Nasdaq dummy equals one if the stock is listed on the Nasdaq exchange and zero otherwise. The Dividend dummy equals one if the firm paid dividends in the prior fiscal year and zero otherwise. The stock price is measured as the closing share price at the end of the previous quarter.

We also control valuation uncertainty, measured by turnover ratio (*Turnover*) and idiosyncratic volatility (*IVol*), which are important drivers of overconfidence as documented by Kumar (2009). *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. For Nasdaq-listed stocks, trading volume is halved to adjust for double counting, following Gao and Ritter (2010). Following Ang et al. (2006), we calculate the monthly stock idiosyncratic volatility as the standard deviation of the daily return residuals in a month based on the Fama and French (1993) three-factor model and then average it in a quarter. Finally, we control the information environment and the market microstructure. Analyst coverage (*ACov*) is measured as the natural logarithm of one plus the average number of analysts covering the firm in a quarter, and institutional holdings (*Inst*) is the natural logarithm of one plus institutional ownership scaled by shares outstanding at quarter-end. To capture microstructure frictions, we include the bid-ask spread, defined as the average daily closing percentage spread in a quarter.[10]

The summary statistics of the variables of interest and the control variables at the firm-quarter level are presented in Table 2. The mean post-trade sell-buy return differential (*PTSBD*) is –0.27 %, indicating that on average, stocks sold by retail investors underperform those they purchase by 0.27 % in the next 30 days, indicating that retail investors, on average, achieve modest gains in our sample. The average *Beta* is 1.42, indicating that retail investors tend to trade high beta stocks. The average firm size, measured as the natural logarithm of market capitalization in millions of dollars, is 5.16, slightly higher than the typical value for stocks listed on major U.S. exchanges during the same period (with an average firm size of about 4.5). For momentum, the mean and median percentage returns for the preceding 11 months are both 7 %. The average log book-to-market is –0.89, which implies that market values generally exceed book values for the sample stocks. The average and median stock prices are 17.00 and 14.00, respectively. 33 % of the firms in the sample pay dividends and 60 % are listed on the Nasdaq exchange. The mean turnover ratio is 0.43, indicating that 0.43 % of the firm's outstanding shares change hands daily on average. Institutional investors hold an average of 8 % of shares, and the distribution of institutional holding is positively skewed. The average analyst coverage is 1.21 in logs, equivalent to about 2.35 analysts per stock. The average bid-ask spread is 3.34 %, and the mean idiosyncratic volatility is 2.97 % per day.

We also report two variables that are not part of the control set, but are used in later analyses. Trading volume (*TV*) is measured as the natural logarithm of one plus the sum of shares traded (in thousands) by retail investors per quarter for a stock. The mean *TV* is

[10] We exclude daily bid-ask spreads in the top 1% of the distribution to mitigate potential measurement errors. Since we average daily bid-ask spreads in a quarter for a given firm, this exclusion has little impact on the number of observations in our dataset.

1.53, corresponding to about 3620 retail-traded shares per quarter. This is small relative to total volume because it only captures trades made by retail investors. *TV* is used as a proxy for disagreement among retail investors. Firm age, measured as the number of years since first appearing in CRSP, averages 11.39, and is later used as a proxy for information uncertainty.

4. Empirical analysis

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

4.1. Main results

To estimate the causal impact of EDGAR on overconfidence, we use the stacked diff-in-diff analysis following Cengiz et al. (2019) and Baker et al. (2022). This method addresses concerns about heterogeneity of treatment effects across staggered implementations by constructing a "clean" 2 × 2 dataset for each implementation phase. In each dataset, the treatment firms are treated in a certain phase, while the control firms are not yet treated at the end of the relevant window (Cengiz et al., 2019). For each phase of EDGAR implementation, we define the treated firms in the event window [−4, 4], where quarter 0 represents the quarter of EDGAR implementation. For example, if the effective treatment date is 1994Q1, the window begins in 1993Q1 and ends in 1995Q1. In this case, the control groups are composed of firms in phases 7 through 10 as the earliest effective treatment date for phase 7 (1995Q2) occurs after the end of the event window (1995Q1).

Chang et al. (2022) find that firm size is related to assignments of firms to each phase. We follow their approach and apply the nearest-neighbor propensity score matching based on equity market capitalization (in both levels and logs). We construct event samples for phases 2 through 6 since the final four phases lack valid control groups: by design, a control firm must not be treated within the event window.[11] This stacked design mitigates the bias from overlapping treatment timing and supports a more credible identification of the EDGAR effect on investor overconfidence.

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

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

where $PTSBD_{i,t,p}$ captures the post-trading performance of stock **i** in quarter **t** in the implementation phase **p**. A larger (more positive) $PTSBD_{i,t,p}$ indicates that returns earned in 30 days after the sell trades are higher than those following purchase trades. We expect that on average, investor overconfidence—as measured by *PTSBD*, will decline following the implementation of EDGAR.

Post is the treatment indicator that equals one within four quarters after a firm joins EDGAR and zero otherwise. *Post* is undefined in the implementation quarter ($t = 0$).[12] *X* represents a set of control variables. $\gamma_{i,p}$ and $\rho_{t,p}$ denote the firm and year-quarter fixed effects for the implementation phase (p), which are defined as the interactions between the phase and the firm dummies (i) and between the phase and year-quarter dummies (t), respectively. We cluster the standard error at the firm level because the time dimension of our panel data is low (Petersen, 2009).

Table 3 presents the regression results. In column (1), the coefficient of *Post* is −1.233, with a *t*-statistic of −2.87, indicating a significant decline in overconfidence following the implementation of EDGAR. Given that the sample mean and standard deviation of *PTSBD* are –0.27 % and 8.47 %, respectively, the coefficient of −1.233 represents approximately 4.57 times the mean and 14.56 % of the standard deviation. After including control variables that may affect investor overconfidence, the coefficient remains virtually unchanged at −1.249, as shown in column (2), and remains significant at the 1 % level (*t*-stat = −2.89). Overall, these findings indicate that the 30-day post-trade returns following sell trades decrease compared to those following purchase trades after EDGAR is implemented, suggesting a decrease in the overconfidence of retail investors.

4.2. Parallel trends and dynamic effects

An important assumption for the diff-in-diff analysis is that the treated and control groups follow parallel trends in the absence of the EDGAR implementation. To assess the validity of this assumption, we test whether the treated and control firms share similar levels of overconfidence before treatment using our stacked sample. For each implementation phase, we use the same event window [−4, 4] and set the first quarter in each window, i.e., $t = -4$, as the base quarter. *EDGAR* (−3) (*EDGAR* (−2), *EDGAR* (−1)) is an indicator variable that equals one if firms are scheduled to join the EDGAR system in three (two, one) quarters and zero otherwise. *EDGAR* (0) equals one if a firm implements EDGAR this quarter and zero otherwise. *EDGAR* (+1) (*EDGAR* (+2)) is an indicator variable that equals one if firms have joined the EDGAR system one (two) quarter(s) ago and zero otherwise. *EDGAR* (3+) is an indicator variable that

[11] For the event window for phase 7 ending in 1996Q2, the eligible control firms must have implementation dates in 1996Q3 or later. Since the final phase (phase 10) was implemented in 1996Q2, firms in phases 7 through 10 lack valid control groups and are excluded from the stacked diff-in-diff analysis.

[12] The implementation quarter includes trading days both before and after the shock date, blending pre- and post-treatment effects in that quarter.

Table 3
Effect of EDGAR implementation on overconfidence.

	(1)	(2)
Post	−1.233***	−1.249***
	(−2.87)	(−2.89)
Beta		0.233
		(0.48)
Size		0.691
		(0.87)
Mom		0.048
		(0.10)
B/M		0.538
		(1.20)
Price		0.027
		(0.88)
Turnover		−0.001
		(−0.01)
Dividend dummy		1.063
		(1.41)
Nasdaq dummy		−1.014
		(−0.87)
Inst		−2.438
		(−1.16)
ACov		−0.231
		(−0.47)
Bid-ask spread		−0.101
		(−0.51)
IVol		0.095
		(0.23)
Firm FE × Phase FE	Yes	Yes
Year-quarter FE × Phase FE	Yes	Yes
Observations	9249	9249
Adj R-squared	0.016	0.018

This table reports coefficients and *t*-statistics of stacked diff-in-diff regressions of overconfidence on *Post* as Eq. (1). The dependent variable is post-trade sell-buy return differential (*PTSBD*), defined as the quarterly mean of the difference between the 30-day buy-and-hold returns following all sell trades and those following all buy trades for a given stock. *Post* is the treatment indicator that equals one (zero) within four quarters after (before) a firm joins EDGAR, i.e., shock date, and is undefined within the event quarter. The definitions of other explanatory variables are described in Table 2. Standard errors are clustered at the stock level. The *t*-statistics are reported in parentheses. ***, **, and *indicate statistical significance at the 1 %, 5 %, and 10 % levels, respectively.

equals one if firms adopted the EDGAR three or more quarters ago, and zero otherwise. In columns (1) and (2) of Table 4, the coefficients of *EDGAR* (−3), *EDGAR* (−2) and *EDGAR* (−1) are statistically nonsignificant, indicating that the treated and control firms share a common trend. We also observe a statistically significant decrease in overconfidence from the shock quarter and thereafter within the event window, consistent with the main findings.

Fig. 1 illustrates the dynamic effects of EDGAR implementation. We find no significant differences in overconfidence between treated and control firms prior to the implementation, supporting the assumption of parallel trends. There is a marginal effect during the implantation quarter because it includes both days before and after the shock dates. The significant decrease in overconfidence for treated firms occurs after the implementation quarter, and the decrease appears to be permanent.

4.3. Heterogeneity in the effects of EDGAR implementation

To better understand the mechanisms through which EDGAR reduces the overconfidence of retail investors, we reexamine our main analysis using subsamples formed by various measures of firm information uncertainty. As EDGAR implementation decreases information acquisition costs and improves the availability of firms' financial data, the decrease in overconfidence after the inclusion of a firm in EDGAR can be attributed to the increased availability of fundamental information about companies that retail investors use for stock valuation. The effect of the availability of fundamental information is expected to be more pronounced for firms with greater information uncertainty and more difficult to value.

We use four proxies for firm information uncertainty, including firm age (*Age*), book-to-market equity (*B/M*), idiosyncratic volatility (*IVol*), and an indicator of high-tech firms. Firm age, defined as the number of years since the CRSP first covered the firm, is used as the inverse proxy for uncertainty, consistent with evidence that longer listing histories are associated with greater information availability (Barry and Brown, 1985; Leary and Roberts, 2010; Maskara and Mullineaux, 2011). Book-to-market captures the growth and value dimension. Theory and evidence suggest that growth firms disclose less and are endogenously more opaque than value firms,

Table 4
Parallel trends test and dynamic effects.

	(1)	(2)
EDGAR (−3)	−0.990	−1.070
	(−1.15)	(−1.24)
EDGAR (−2)	−0.988	−0.964
	(−1.18)	(−1.16)
EDGAR (−1)	−0.577	−0.585
	(−0.62)	(−0.65)
EDGAR (0)	−1.410*	−1.428*
	(−1.73)	(−1.74)
EDGAR (+1)	−1.728*	−1.764*
	(−1.74)	(−1.86)
EDGAR (+2)	−1.768*	−1.670*
	(−1.91)	(−1.77)
EDGAR (3+)	−1.994**	−1.870**
	(−2.47)	(−2.25)
Controls	No	Yes
Firm FE × Phase FE	Yes	Yes
Year-quarter FE × Phase FE	Yes	Yes
Observations	10,475	10,475
Adj R-squared	0.013	0.015

This table shows the dynamic effects of the EDGAR shock on overconfidence estimated from stacked diff-in-diff regressions. The dependent variable is post-trade sell-buy return differential (*PTSBD*), defined as the quarterly mean of the difference between the 30-day buy-and-hold returns following all sell trades and those following all buy trades for a given stock. *EDGAR* (−3) (*EDGAR* (−2), *EDGAR* (−1)) is an indicator variable that equals one if firms are scheduled to join the EDGAR system in three (two, one) quarter(s), and zero otherwise. *EDGAR* (0) equals one if a firm implements EDGAR this quarter, and zero otherwise. *EDGAR* (+1) (*EDGAR* (+2)) is an indicator variable that equals one if firms have joined the EDGAR system one (two) quarter(s) ago, and zero otherwise. *EDGAR* (3+) is an indicator variable that equals one if firms have adopted the EDGAR three or more quarters ago, and zero otherwise. Control variables are the same as those in Table 3. Standard errors are clustered at the stock level. The *t*-statistics are reported in parentheses. ***, **, and *indicate statistical significance at the 1 %, 5 %, and 10 % levels, respectively.

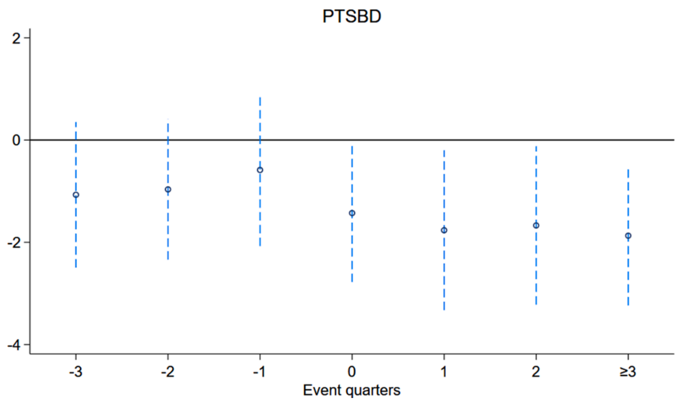

Fig. 1. Dynamic treatment effects. This figure plots the diff-in-diff estimates and their 90 percent confidence intervals of the regression reported in Table 4 column (2). The dependent variable is post-trade sell-buy return differential (*PTSBD*), defined as the quarterly mean of the difference between the 30-day buy-and-hold returns following all sell trades and those following all buy trades for a given stock. The treated firms are included in EDGAR (shock date) in quarter 0. (−3), (−2), and (−1) indicates three, two, and one-quarter(s) before the inclusion, and (1), (2), and (≥3) indicates one, two, and three and more quarter(s) after the inclusion, respectively.

implying higher uncertainty for low book-to-market stocks (Gao and Liang, 2013). Stocks with higher idiosyncratic volatility are often considered to have higher uncertainty and are more difficult to value (Kumar, 2009).

We also classify firms by whether they are listed on the Nasdaq exchange. Firms listed on the Nasdaq exchange tend to be high-tech companies with high information complexity and are difficult to value, due to the specialized industry knowledge required for ordinary retail investors (Chen et al., 2025). We acknowledge potential limitations of this proxy in our sample period. During the 1990s, Nasdaq-listed firms differed from NYSE/AMEX firms not only in industry composition but also in market microstructure and trading environment, which may affect trading outcomes independently of information complexity.[13]

We separately split our sample into subsamples based on each of these four proxies available prior to the beginning of each event window. Firms below (above) the median of age, book-to-market equity, and idiosyncratic volatility are young (old), growth (value), and low (high) IVol firms, respectively. Firms listed on the Nasdaq exchange are classified as high-tech firms, and the remaining are considered as low-tech firms. If the implementation of EDGAR reduces retail investor overconfidence by increasing the availability of firm fundamental information, we expect stronger EDGAR effects among young firms, low book-to-market firms, high-IVol firms, and high-tech firms.

Table 5 shows the results of the stacked diff-in-diff analysis for sub-samples. In columns (1) and (2) of Panel A, the coefficients of *Post* are −2.099 and −2.101 with the *t*-statistics of −3.17 and −3.13, respectively, indicating that for young firms, *PTSBD* decreases by around 2.10 % after the inclusion of EDGAR. On the contrary, this effect is smaller and not significant for old firms, as shown in columns (3) and (4). The differences between the coefficients for young and old firms are statistically significant at a 10 % level. Panel B presents results for sub-samples formed by the book-to-market equity ratio. We observe that the decrease in overconfidence is more pronounced for growth firms than for value firms. Without control variables, *PTSBD* declines by a statistically significant 1.66 %, compared to only a 0.34 % and nonsignificant decline for value firms. The difference between these coefficients is also statistically significant. The results remain qualitatively the same after adding control variables in the regressions. Panel C shows consistent results when we use idiosyncratic volatility as a measure of information uncertainty. *PTSBD* decreases by a statistically significant 2.23 % for firms with high idiosyncratic volatility in column (1) and by a nonsignificant 0.486 % for firms with low idiosyncratic volatility in column (3). The difference in coefficients is highly and statistically significant. The results are robust to including control variables, as shown in columns (2) and (4). Panel D shows that without control variables, *PTSBD* decreases by 1.915 % (*t*-stat = −3.17) for high-tech firms but only decreases by 0.727 % (*t*-stat = −1.00) for low-tech firms, and such results hold qualitatively when we add controls. As discussed earlier, we interpret the results based on the Nasdaq dummy with caution since Nasdaq listing may also reflect differences in market microstructure during our sample period. Nevertheless, our conclusions do not rely on this proxy alone and are supported by the other three heterogeneity measures as discussed above.

In conclusion, our analyses show that the implementation of EDGAR significantly reduces overconfidence of retail investors, particularly for trading firms with high information uncertainty. This decrease can be attributed to the improved information environment and the reduced costs that investors face when acquiring fundamental information about these firms.

5. Overconfidence, disagreement, and mispricing

In this section, we examine the implications of the reduced overconfidence of retail investors following the implementation of EDGAR for their disagreement and stock mispricing in general.

5.1. Overconfidence and disagreement

Psychological and experimental research suggests that individuals tend to overweight confirmatory signals and underweight contradictory evidence, leading to overconfidence that exceeds actual accuracy (Einhorn, 1980; Griffin and Tversky, 1992). Theoretically, these tendencies, combined with limits to arbitrage, can generate persistent disagreement and asset-price bubbles (Scheinkman and Xiong, 2003; Xiong, 2013). Empirically, Chang et al. (2022) find that the improved information disclosure through EDGAR reduces investor disagreement, consistent with the argument that a more transparent information environment disciplines belief formation. Using the setting of EDGAR implementation, we examine whether the decrease in investor overconfidence is associated with a decrease in retail investor disagreement.

We employ a two-stage least squares (2SLS) analysis on the same stacked sample that we examined previously. In the first stage, we regress *PTSBD*, the overconfidence measure, on the EDGAR *Post* and controls with fixed effects, as in Eq. (1). In the second stage, we regress the disagreement measures on the instrumented *PTSBD*, $\widehat{PTSBD}$, to examine whether the reduction in disagreement can be attributed to the decline in retail investor overconfidence following the EDGAR. Following Chang et al. (2022), we use the retail share trading volume as the measure of disagreement. Column (1) of Table 6 presents the result of the first stage of the 2SLS regression (which is the same regression with the control variables reported in Table 3). The result shows a decrease in overconfidence after the inclusion of EDGAR.

Column (2) of Table 6 shows that the coefficient of the *PTSBD* instrumented by EDGAR in the second-stage regression is positive and statistically significant, consistent with the view that lower retail overconfidence contributes to the reduction in retail trading volume after the implementation of EDGAR.[14] The cluster-robust Kleibergen–Paap Wald F-statistic is 8.32, slightly less than the rule-of-thumb threshold of 10, suggesting a potential problem of weak instruments. We further conduct the Anderson–Rubin test with $\chi_1^2 = 20.16$ ($p < 0.001$), rejecting the null that the coefficient of the endogenous regressor is zero. Therefore, we conclude that the second-stage inference is reliable and not driven by weak-instrument problems.

We acknowledge that the implementation of EDGAR may also affect trading volume through channels other than overconfidence,

[13] Nasdaq was historically a multi-dealer, quote-driven market, whereas NYSE relied more on an auction/specialist structure. The execution cost on the Nasdaq was higher than NYSE in our sample period.

[14] The evidence is robust when we use the dollar volume as a proxy of disagreement.

Table 5
Effect of EDGAR implementation on overconfidence: Subsample analysis.

Panel A	Young firms		Old firms	
	(1)	(2)	(3)	(4)
Post	−2.099***	−2.101***	−0.560	−0.430
	(−3.17)	(−3.13)	(−0.99)	(−0.75)
Controls	No	Yes	No	Yes
Firm FE × Phase FE	Yes	Yes	Yes	Yes
Year-quarter FE × Phase FE	Yes	Yes	Yes	Yes
Observations	4445	4445	4804	4804
Adj R-squared	0.023	0.026	0.006	0.012
Diff between column (n) and ($n+2$)	−1.539	−1.671		
p-value	0.10	0.04		

Panel B	Growth firms		Value firms	
	(1)	(2)	(3)	(4)
Post	−1.663**	−1.471**	−0.344	−0.360
	(−2.52)	(−2.36)	(−0.48)	(−0.49)
Controls	No	Yes	No	Yes
Firm FE × Phase FE	Yes	Yes	Yes	Yes
Year-quarter FE × Phase FE	Yes	Yes	Yes	Yes
Observations	4957	4957	3364	3364
Adj R-squared	0.004	0.003	0.039	0.047
Diff between column (n) and ($n+2$)	−1.319	−1.110		
p-value	0.07	0.08		

Panel C	High IVol firms		Low IVol firms	
	(1)	(2)	(3)	(4)
Post	−2.229***	−2.126***	−0.486	−0.567
	(−2.98)	(−2.80)	(−0.99)	(−1.14)
Controls	No	Yes	No	Yes
Firm FE × Phase FE	Yes	Yes	Yes	Yes
Year-quarter FE × Phase FE	Yes	Yes	Yes	Yes
Observations	4326	4326	4908	4908
Adj R-squared	0.012	0.013	0.023	0.026
Diff between column (n) and ($n+2$)	−1.743	−1.559		
p-value	0.01	0.03		

Panel D	High-tech firms		Low-tech firms	
	(1)	(2)	(3)	(4)
Post	−1.915***	−1.964***	−0.727	−0.660
	(−3.17)	(−3.24)	(−1.00)	(−0.98)
Controls	No	Yes	No	Yes
Firm FE × Phase FE	Yes	Yes	Yes	Yes
Year-quarter FE × Phase FE	Yes	Yes	Yes	Yes
Observations	5549	5549	3675	3675
Adj R-squared	0.015	0.018	0.004	0.011
Diff between column (n) and ($n+2$)	−1.188	−1.304		
p-value	0.03	0.04		

This table reports the coefficients and *t*-statistics of stacked diff-in-diff regressions of overconfidence on *Post*. The dependent variable is post-trade sell-buy return differential (*PTSBD*), defined as the quarterly mean of the difference between the 30-day buy-and-hold returns following all sell trades and those following all buy trades for a given stock. *Post* is the treatment indicator that equals one (zero) within four quarters after (before) a firm joins EDGAR, i.e., shock date, and is undefined within the event quarter. Panel A shows the results of young vs. old firms. Young (Old) firms are defined as firms whose age is lower (higher) than the median. Panel B shows the results of growth vs. value firms. Growth (Value) firms are defined as firms whose *B/M* is lower (higher) than the median. Panel C shows the results of subsamples split by the median of idiosyncratic volatility (IVol). Panel D shows the results of high-tech vs. low-tech firms. High-tech firms are those listed on Nasdaq, and low-tech firms are all others. Control variables are the same as those in Table 3. The differences between the coefficients of *Post* for the two subsamples and the associated *p*-values are reported at the bottom in each panel. Standard errors are clustered at the stock level. The *t*-statistics are reported in parentheses. ***, **, and *indicate statistical significance at the 1 %, 5 %, and 10 % levels, respectively.

including improved liquidity, lower participation costs, or changes in investor base. While we cannot fully rule out all unobserved channels, we mitigate these concerns by controlling for liquidity proxies (bid–ask spread and turnover) and for changes in the information environment (analyst coverage and institutional ownership), and by restricting our sample to the investors who have traded

Table 6
Overconfidence and disagreement.

	PTSBD (1)	*TV* (2)
Post	−1.249***	
	(−2.89)	
$\widehat{PTSBD}$		0.167***
		(2.36)
Controls	Yes	Yes
Firm FE × Phase FE	Yes	Yes
Year-quarter FE × Phase FE	Yes	Yes
Observations	9249	9249
Adj R-squared	0.018	0.62

This table reports 2SLS regression results of the impacts of overconfidence on disagreement. Post-trade sell-buy return differential (*PTSBD*) is the quarterly mean of the difference between the 30-day buy-and-hold returns following all sell trades and those following all buy trades for a given stock. Disagreement is measured by retail investors' trading volume (*TV*), defined as the natural logarithm of one plus the sum of shares traded (in thousands) in a quarter for a stock. *Post* is the treatment indicator that equals one (zero) within four quarters after (before) a firm joins EDGAR, i.e., shock date, and is undefined within the event quarter. (1) is the first-stage regression and (2) is the second-stage regression. Control variables are the same as those in Table 3. Standard errors are clustered at the stock level. The *t*-statistics are reported in parentheses. ***, **, and *indicate statistical significance at the 1 %, 5 %, and 10 % levels, respectively.

stocks prior to EDGAR implementation to maintain the same investor base around the event.

5.2. Overconfidence and mispricing

Overconfident investors who overweight their private signals and recent confirming signals may push prices away from fundamentals when arbitrage is limited (Daniel et al., 1998; Shleifer and Vishny, 1997). Hirshleifer and Ma (2024) show that the implementation of EDGAR reduces stock mispricing, suggesting that better information environments discipline speculative trading behaviors. We examine whether the implementation of EDGAR reduces overconfidence-driven mispricing.

We use the following regression specification

$$\begin{aligned} Ret_{i,t+1,p} = {} & \beta_1 \times \left(Misp_{i,t,p} \times High\_PTSBD_{i,t,p} \times Post_{i,t,p}\right) + \beta_2 \times Misp_{i,t,p} + \beta_3 \times High\_PTSBD_{i,t,p} + \beta_4 \times Post_{i,t,p} + \beta_5 \\ & \times \left(Misp_{i,t,p} \times Post_{i,t,p}\right) + \beta_6 \times \left(Misp_{i,t,p} \times High\_PTSBD_{i,t,p}\right) + \beta_7 \times \left(High\_PTSBD_{i,t,p} \times Post_{i,t,p}\right) + \beta_8 \times X_{i,t,p} + \rho_{t,p} \\ & + \epsilon_{i,t+1,p}, \end{aligned} \tag{2}$$

where $Ret_{i,t+1,p}$ is the three month cumulative return of stock i in quarter $t+1$ in implementation phase p.[15] $Misp_{i,t,p}$ is the mispricing signal of stock i at the end of quarter t. We use the negative of the mispricing signal constructed by Stambaugh and Yuan (2017) so that a larger $Misp_{i,t,p}$ reflects greater underpricing of stock i at the end of quarter t. $High\_PTSBD_{i,t,p}$ is the indicator variable which equals one if *PTSBD* is above the 80th percentile of the distribution for stock i in quarter t, and zero otherwise. $X_{i,t,p}$ represents control variables, including market beta, size, momentum, and book-to-market ratio, defined in Section 3.3. $\rho_{t,p}$ represents the year-quarter fixed effect for the implementation phase (p), defined as the interactions between the implementation phase and the year-quarter dummies. If EDGAR improves the information environment and reduces investor overconfidence, the overconfidence-driven component of mispricing should decline following the inclusion of EDGAR.

Table 7 presents the regression results of Eq. (2). Across all columns, the coefficients of *Misp* are positive and statistically significant, suggesting that the mispricing signal predicts cross-sectional stock returns in our sample. In column (1), the coeffeient of *Misp* is 0.134, indicating that stocks with one standard deviation higher misprcing signal earn 1.85 % more in the next quarter.[16] In column (2), the significantly positive coefficient of the interaction term *Misp* × *High_PTSBD* is evidence of overconfidence-driven mispricing, i.e., the predictive power of the mispricing signal is greater when the overconfidence of retail investors is high. The coefficient of this interaction term is 0.152, indicating that among stocks subject to high overconfidence of retail investors, one standard deviation variation in the mispricing signal corresponds to an additional 2.10 % variation in quarterly returns. In column (3), the coefficient of the triple interaction term *Misp* × *High_PTSBD* × *Post* is negative, suggesting that the implementation of EDGAR reduces the overconfidence-

[15] Because *PTSBD* is calculated as the 30-day returns after trades in quarter t, i.e., *PTSBD* can be realized in the first month of the next quarter, we define $Ret_{i,t+1}$ as the three-month cumulative return starting from the second month after quarter t to avoid potential overlapping between *PTSBD* and the dependent variable *Ret*.

[16] The mean and standard deviation of *Misp* are −51.67 and 13.81, respectively.

Table 7
Overconfidence-driven mispricing.

	(1)	(2)	(3)
Misp	0.134***	0.106***	0.097***
	(4.91)	(3.58)	(2.59)
Misp × *High_PTSBD*		0.152**	0.170**
		(2.31)	(2.26)
Misp × *High_PTSBD* × *Post*			−0.165
			(−1.14)
High_PTSBD		8.859**	8.944**
		(2.50)	(2.16)
Post			1.124
			(0.36)
Post × *High_PTSBD*			0.066
			(1.15)
Post × *Misp*			−4.478
			(−0.57)
Beta	−0.012	−0.028	−0.116
	(−0.03)	(−0.06)	(−0.25)
Size	−0.279	−0.243	−0.158
	(−0.77)	(−0.67)	(−0.43)
Mom	1.049	1.039	0.957
	(1.01)	(1.00)	(0.92)
B/M	0.893	0.876	1.046*
	(1.58)	(1.57)	(1.84)
Year-quarter FE × Phase FE	Yes	Yes	Yes
Observations	7961	7961	7961
Adj R-squared	0.095	0.096	0.097

This table reports the coefficients and *t*-statistic of the regressions of the future three-month cumulative stock returns as Eq. (2). *Misp* is the mispricing measure from Stambaugh and Yuan (2017) multiplied by −1. Stocks with the highest (least negative) values of *Misp* are the most "underpriced". *High_PTSBD* is an indicator variable which equals one if a stock's *PTSBD* is above the 80th percentile of the distribution and zero otherwise. Post-trade sell-buy return differential (*PTSBD*) is the quarterly mean of the difference between the 30-day buy-and-hold returns following all sell trades and those following all buy trades for a given stock. *Post* is the treatment indicator that equals one (zero) within four quarters after (before) a firm joins EDGAR, i.e., shock date, and is undefined within the event quarter. The definitions of other explanatory variables are described in Table 2. Standard errors are clustered at the stock level. The *t*-statistics are reported in parentheses. ***, **, and *indicate statistical significance at the 1 %, 5 %, and 10 % levels, respectively.

driven mispricing. The coefficient is −0.165, not statistically significant; however, it is of a similar magnitude to the interaction term *Misp* × *High_PTSBD* (0.17). Overall, the results are consistent with the implementation of EDGAR mitigating pre-EDGAR overconfidence-driven mispricing in an economically meaningful way.

6. Robustness tests

We perform several robust checks on the main findings in the paper. As we discussed earlier, the implementation dates of the first four phases of EDGAR in 1993 as shown in Table 1 are earlier than the online date (January 17, 1994) when the EDGAR filings became freely accessible to investors. Because retail investors still faced substantial access fees before the online date, the effects of the EDGAR implementation on overconfidence of retail investors are unlikely to occur after the implementation dates of these phases. We re-estimate the stacked diff-in-diff regressions as in Eq. (1) with the actual implementation dates for phases of 2, 3, and 4 of EDGAR implementation as a falsification test.[17] Table 8 shows that the coefficient of *Post* is nonsignificant, indicating no significant decrease in overconfidence after this placebo timing of the EDGAR implementation. This result not only supports the validity of our stacked diff-in-diff analysis to identify the effect of EDGAR implementation on retail investor overconfidence, but also highlights the importance of freely accessible information in this setting.

To address concerns that our main results may be driven by specific phases of the EDGAR implementation, we perform a diff-in-diff analysis for each phase with firm and year-quarter fixed effects. Table 9 shows that the coefficient of *Post* is negative in every phase except for phase 3 with control variables, where the positive coefficient is not statistically significant. The coefficients are negative and statistically significant for phases 2 and 6. Overall, these results indicate that the observed decreases in retail investor overconfidence after the implementation of EDGAR are not driven by a specific implementation phase.

[17] The authors thank the referee for suggesting this falsification test.

Table 8
Falsification test.

	(1)	(2)
Post	−0.339	0.016
	(−0.83)	(0.04)
Controls	No	Yes
Firm FE × Phase FE	Yes	Yes
Year-quarter FE × Phase FE	Yes	Yes
Observations	5440	5440
Adj R-squared	0.006	0.010

This table reports coefficients and *t*-statistics of stacked diff-in-diff regressions of overconfidence on *Post* as Eq. (1). The dependent variable is post-trade sell-buy return differential (*PTSBD*), defined as the quarterly mean of the difference between the 30-day buy-and-hold returns following all sell trades and those following all buy trades for a given stock. *Post* is re-defined based on the implementation date in Table 1. *Post* equals one within four quarters after a firm joins EDGAR (implementation date) and zero otherwise, and is undefined within the event quarter. Only firms in the phases 2–4 are included in the sample. Control variables are the same as those in Table 3. Standard errors are clustered at the stock level. The *t*-statistics are reported in parentheses. ***, **, and *indicate statistical significance at the 1 %, 5 %, and 10 % levels, respectively.

Table 9
Effects of EDGAR implementation on overconfidence: By each implementation phase.

	(1)	(2)
Phase 2	−2.433**	−2.256*
	(−2.05)	(−1.94)
Phase 3	−0.343	0.193
	(−0.40)	(0.21)
Phase 4	−0.671	−0.504
	(−0.97)	(−0.72)
Phase 5	−0.940	−0.876
	(−1.35)	(−1.28)
Phase 6	−2.565***	−2.870***
	(−2.88)	(−3.16)
Firm FE	Yes	Yes
Year-quarter FE	Yes	Yes
Controls	No	Yes

The table reports the coefficients and *t*-statistics of the diff-in-diff regressions of overconfidence on *Post* for each implantation phase. The dependent variable is post-trade sell-buy return differential (*PTSBD*), defined as the quarterly mean of the difference between the 30-day buy-and-hold returns following all sell trades and those following all buy trades for a given stock. 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. *Post* is the treatment indicator that equals one (zero) within four quarters after (before) a firm joins EDGAR, i.e., shock date, and is undefined within the event quarter. Control variables are the same as those in Table 3. Standard errors are clustered at the stock level. The *t*-statistics are reported in parentheses. ***, **, and *indicate statistical significance at the 1 %, 5 %, and 10 % levels, respectively.

7. Conclusion

We examine whether information acquisition costs affect the overconfidence of retail investors. Using the stacked diff-in-diff analysis and the staggered implementation of the EDGAR system as an exogenous shock to the cost of accessing firm disclosures, we find that improved information access leads to a significant reduction in retail investor overconfidence, measured by post-trade performance based on transaction data from a large U.S. brokerage. We also find that the decrease in overconfidence is more pronounced for firms with high information uncertainty. These include firms that are younger, exhibit stronger growth characteristics, experience greater stock-specific volatility, or are listed on the Nasdaq exchange. These results support the view that the decline in overconfidence after the implementation of EDGAR is largely driven by a reduction in information uncertainty.

We also examine the broader market implications of this decrease in overconfidence. We document that the decrease in the overconfidence of retail investors after the implementation of EDGAR is associated with a reduction in investor disagreement and a weakening of overconfidence-driven stock mispricing. This study contributes to the literature by offering new evidence that the information environment plays a critical role in influencing investor behavioral biases and stock market efficiency.

CRediT authorship contribution statement

Gang Li: Writing – review & editing, Supervision, Methodology, Formal analysis, Conceptualization. **Shuqi Wang:** Writing – review & editing, Writing – original draft, Visualization, Methodology, Formal analysis, Data curation, Conceptualization. **K.C. John Wei:** Writing – review & editing, Supervision, Methodology, Funding acquisition, Conceptualization.

References

Adam, Tim R., Fernando, Chitru S., Golubeva, Evgenia, 2015. Managerial overconfidence and corporate risk management. J. Bank Financ. 60, 195–208.
An, Li, 2016. Asset pricing when traders sell extreme winners and losers. Rev. Financ. Stud. 29 (3), 823–861.
Ang, Andrew, Hodrick, Robert J., Xing, Yuhang, Zhang, Xiaoyan, 2006. The cross-section of volatility and expected returns. J. Finance 61 (1), 259–299.
Baker, Andrew C., Larcker, David F., Wang, Charles C.Y., 2022. How much should we trust staggered difference-in-differences estimates? J. Financ. Econ. 144 (2), 370–395.
Barber, Brad M., Odean, Terrance, 2000. Trading is hazardous to your wealth: the common stock investment performance of individual investors. J. Finance 55 (2), 773–806.
Barber, Brad M., Odean, Terrance, 2001. Boys will be boys: gender, overconfidence, and common stock investment. Q. J. Econ. 116 (1), 261–292.
Barber, Brad M., Odean, Terrance, 2002. Online investors: do the slow die first? Rev. Financ. Stud. 15 (2), 455–487.
Barry, Christopher B., Brown, Stephen J., 1985. Differential information and security market equilibrium. J. Financ. Quant. Anal. 20 (4), 407–422.
Ben-David, Itzhak, Hirshleifer, David, 2012. Are investors really reluctant to realize their losses? Trading responses to past returns and the disposition effect. Rev. Financ. Stud. 25 (8), 2485–2532.
Blankespoor, Elizabeth, deHaan, Ed, Marinovic, Ivan, 2020. Disclosure processing costs, investors' information choice, and equity market outcomes: a review. J. Account. Econ. 70 (2–3), 101344.
Biais, Bruno, Hilton, Denis, Mazurier, Karine, Pouget, Sébastien, 2005. Judgemental overconfidence, self-monitoring, and trading performance in an experimental financial market. Rev. Econ. Stud. 72 (2), 287–312.
Cengiz, Doruk, Dube, Arindrajit, Lindner, Attila, Zipperer, Ben, 2019. The effect of minimum wages on low-wage jobs. Q. J. Econ. 134 (3), 1405–1454.
Chang, Yen-Cheng, Hsiao, Pei-Jie, Ljungqvist, Alexander, Tseng, Kevin, 2022. Testing disagreement models. J. Finance 77 (4), 2239–2285.
Chang, Yen-Cheng, Ljungqvist, Alexander, Tseng, Kevin, 2023. Do corporate disclosures constrain strategic analyst behavior? Rev. Financ. Stud. 36 (8), 3163–3212.
Chen, Yangfa, Jiang, Ji, Liu, Jie, Liu, Xiao, Wu, Weili, 2025. Registration system reform, information environment, and market manipulation. J. Corp. Finance 93, 102797.
Chou, Hsin-I, Li, Mingyi, Yin, Xiangkang, Zhao, Jing, 2021. Overconfident institutions and their self-attribution bias: evidence from earnings announcements. J. Financ. Quant. Anal. 56 (5), 1738–1770.
Chui, Andy C.W., Titman, Sheridan, John Wei, K.C., 2010. Individualism and momentum around the world. J. Finance 65 (1), 361–392.
Daniel, Kent, Hirshleifer, David, Subrahmanyam, A., 1998. Investor psychology and security market under- and overreactions. J. Finance 53 (6), 1839–1885.
Daniel, Kent D, Hirshleifer, David, Subrahmanyam, Avanidhar, 2001. Overconfidence, arbitrage, and equilibrium asset pricing. J. Finance 56 (3), 921–965.
De Bondt, Werner F.M., Thaler, Richard H., 1995. Financial decision-making in markets and firms: a behavioral perspective. Handb. Oper. Res. Manag. Sci. 9, 385–410.
Einhorn, Hillel J, 1980. Overconfidence in judgment. New Dir. Methodol. Soc. Behav. Sci. 4 (1), 1–16.
Fama, Eugene F, French, Kenneth R, 1992. The cross-section of expected stock returns. J. Finance 47 (2), 427–465.
Fama, Eugene F, French, Kenneth R., 1993. Common risk factors in the returns on stocks and bonds. J. Financ. Econ. 33 (1), 3–56.
Fischhoff, Baruch, Slovic, Paul, Lichtenstein, Sarah, 1977. Knowing with certainty: the appropriateness of extreme confidence. J. Exp. Psychol.: Hum. Percept. Perform. 3 (4), 552.
Gao, Meng, Huang, Jiekun, 2020. Informing the market: the effect of modern information technologies on information production. Rev. Financ. Stud. 33 (4), 1367–1411.
Gao, Pingyang, Liang, Pierre Jinghong, 2013. Informational feedback, adverse selection, and optimal disclosure policy. J. Account. Res. 51 (5), 1133–1158.
Gao, Xiaohui, Ritter, Jay R., 2010. The marketing of seasoned equity offerings. J. Financ. Econ. 97 (1), 33–52.
Gervais, Simon, Odean, Terrance, 2001. Learning to be overconfident. Rev. Financ. Stud. 14 (1), 1–27.
Glaser, Markus, Weber, Martin, 2007. Overconfidence and trading volume. Geneva Risk Insur. Rev. 32 (1), 1–36.
Goldstein, Itay, Yang, Shijie, Zuo, Luo, 2023. The real effects of modern information technologies: evidence from the EDGAR implementation. J. Account. Res. 61 (5), 1699–1733.
Griffin, Dale, Tversky, Amos, 1992. The weighing of evidence and the determinants of confidence. Cogn. Psychol. 24 (3), 411–435.
Grinblatt, Mark, Han, Bing, 2005. Prospect theory, mental accounting, and momentum. J. Financ. Econ. 78 (2), 311–339.
Hirshleifer, David, 2001. Investor psychology and asset pricing. J. Finance 56 (4), 1533–1597.
Hirshleifer, David, and Liang Ma, 2024. The effect of new information technologies on asset pricing anomalies. Working paper.
Kahneman, Daniel, 2003. Maps of bounded rationality: psychology for behavioral economics. Am. Econ. Rev. 93 (5), 1449–1475.
Kahneman, Daniel, Tversky, Amos, 1973. On the psychology of prediction. Psychol. Rev. 80 (4), 237–251.
Kim, Yong Hyuck, Ivkovich, Zoran, Muravyev, Dmitriy, 2021. Causal Effect of Information Costs On Asset Pricing Anomalies. Working paper available at SSRN 3921785.
Kumar, Alok, 2009. Hard-to-value stocks, behavioral biases, and informed trading. J. Financ. Quant. Anal. 44 (6), 1375–1401.
Kyle, Albert S, Albert Wang, F., 1997. Speculation duopoly with agreement to disagree: can overconfidence survive the market test? J. Finance 52 (5), 2073–2090.
Leary, Mark T., Roberts, Michael R., 2010. The pecking order, debt capacity, and information asymmetry. J. Financ. Econ. 95 (3), 332–355.
Liu, Yibin, 2019. Information Acquisition Costs and misreporting: Evidence from the Implementation of EDGAR. Working paper.
Maskara, Pankaj K., Mullineaux, Donald J., 2011. Information asymmetry and self-selection bias in bank loan announcement studies. J. Financ. Econ. 101 (3), 684–694.
Odean, Terrance, 1998a. Are investors reluctant to realize their losses? J. Finance 53 (5), 1775–1798.
Odean, Terrance, 1998b. Volume, volatility, price, and profit when all traders are above average. J. Finance 53 (6), 1887–1934.
Odean, Terrance, 1999. Do investors trade too much? Am. Econ. Rev. 89 (5), 1279–1298.
Petersen, Mitchell A., 2009. Estimating standard errors in finance panel data sets: comparing approaches. Rev. Financ. Stud. 22 (1), 435–480.
Scheinkman, Jose A., Xiong, Wei, 2003. Overconfidence and speculative bubbles. J. Polit. Econ. 111 (6), 1183–1220.
Shleifer, A., Vishny, R.W., 1997. The limits of arbitrage. J. Finance 52 (1), 35–55.
Stambaugh, Robert F, Yuan, Yu, 2017. Mispricing factors. Rev. Financ. Stud. 30 (4), 1270–1315.
Statman, Meir, Thorley, Steven, Vorkink, Keith, 2006. Investor overconfidence and trading volume. Rev. Financ. Stud. 19 (4), 1531–1565.
Xiong, Wei, 2013. Bubbles, crises, and heterogeneous beliefs. In: Langsam, Joseph A., Fouque, Jean-Pierre (Eds.), Handbook For Systemic Risk. Cambridge University Press, Cambridge, pp. 663–713.
Zhang, X.Frank, 2006. Information uncertainty and stock returns. J. Finance 61 (1), 105–137.