June 28, 2026

When AI Reads the 10-Ks: The Effects of ChatGPT on Trading and Price Dynamics[†]

Toghrul Aghbabali,[a]

Kee Chung[b,*]

Sahn-Wook Huh[c]

Abstract

This study examines how ChatGPT's release reshaped the U.S. equity market by democratizing information processing. We devise a metric, *CSA*, to capture ChatGPT's ability to interpret financial disclosures. Using a difference-in-differences design, we show that high-*CSA* firms exhibit significant declines in informed trading, adverse selection, and high-frequency volatility. We also find increased order imbalances primarily driven by retail investors, signaling their synchronized reactions to public information. These suggest that generative AI narrows cognitive gaps between retail and institutional investors, improving market quality without accessing private information. However, heterogeneity persists across sectors, highlighting regulatory implications for disclosure standards in AI-augmented environments.

JEL Classification: G12; G14; G41
Keywords*:* Coverage Score Average (*CSA*); Large Language Models; Generative AI; ChatGPT; Processing Public Information; Information Asymmetry; Adverse-Selection Cost; Informed Trading; Order Imbalances; Retail Investors

[a] School of Management, University (SUNY) at Buffalo, Buffalo, NY 14260-4000. Email: toghrula@buffalo.edu; Tel: 646-799-3402.

[b] School of Management, University (SUNY) at Buffalo, Buffalo, NY 14260-4000. Email: keechung@buffalo.edu; Tel: 716-645-3262.

[c] School of Management, University (SUNY) at Buffalo, Buffalo, NY 14260-4000. Email: swhuh@buffalo.edu; Tel: 716-645-5435.

[*] *Corresponding Author.*

[†] We are grateful to the UB Department of Finance and the School of Management for their logistical and financial support, which enabled the derivation of our primary metric, Coverage Score Average (*CSA*), using the GPT-3.5 Turbo large language model.

When AI Reads the 10-Ks: The Effects of ChatGPT on Trading and Price Dynamics

Abstract

This study examines how ChatGPT's release reshaped the U.S. equity market by democratizing information processing. We devise a metric, *CSA*, to capture ChatGPT's ability to interpret financial disclosures. Using a difference-in-differences design, we show that high-*CSA* firms exhibit significant declines in informed trading, adverse selection, and high-frequency volatility. We also find increased order imbalances primarily driven by retail investors, signaling their synchronized reactions to public information. These suggest that generative AI narrows cognitive gaps between retail and institutional investors, improving market quality without accessing private information. However, heterogeneity persists across sectors, highlighting regulatory implications for disclosure standards in AI-augmented environments.

JEL Classification: G12; G14; G41

Keywords*:* Coverage Score Average (*CSA*); Large Language Models; Generative AI; ChatGPT; Processing Public Information; Information Asymmetry; Adverse-Selection Cost; Informed Trading; Order Imbalances; Retail Investors

1. Introduction

Information is at the heart of financial markets, and how well markets incorporate it largely depends on investors' ability to understand it. Although the Internet and regulatory reforms have made corporate information more accessible, significant differences still exist in investors' capacity to interpret it. Institutional investors use professional analysts, data scientists, and computational tools to analyze complex disclosures, while most retail investors lack the training and resources to do so. These differences in information-processing skills are a key feature of modern financial markets: even when investors access the same filings and news, they do not interpret them in the same way. Whether new technologies can help close this gap is central to research on fair and efficient markets. Our study aligns with the findings of Hong and Stein (2007) and Tetlock (2007), which indicate that information asymmetry can result from private signals or a better understanding of public information.

The rise of large language models (LLMs), like ChatGPT, could be revolutionary. For the first time, accessible AI can translate complex financial statements into simple language, summarize corporate filings, and provide clear explanations of a company's performance. Unlike earlier innovations such as electronic disclosure systems or algorithmic trading, LLMs can democratize understanding rather than just grant access. While GPT-based models existed before November 2022, their use was mostly limited to technical users and developers. The public launch of ChatGPT on November 30, 2022 significantly increased accessibility and awareness by offering a free, easy-to-use interface. This greatly lowered the costs of adoption and experimentation, made LLM capabilities more visible, and promptly influenced perceptions of the economic value of generative AI. Public and investor interest surged, with ChatGPT reaching around 100 million users within months (Reuters, 2023). By significantly reducing the cost of understanding financial information, this release provided an unprecedented opportunity for retail investors to gain the same insights as professionals — raising an important question for research on fair and efficient markets: can this technological shift help close information gaps and improve market quality?

In this paper, we explore whether the emergence of generative AI has narrowed the informational gap between retail and institutional investors. Specifically, we examine whether the public release of ChatGPT changed the level of information asymmetry in the U.S. equity market. To do this, we introduce a new firm-level metric that measures ChatGPT's ability to read and interpret firms' financial statements. Using this metric within a difference-in-differences (DiD) framework centered on ChatGPT's launch, we investigate whether firms most "coverable" by ChatGPT experience a reduction in information asymmetry among traders. Our study starts from the premise that ChatGPT may have caused a different type of information shock. Unlike previous technological or regulatory changes [e.g., EDGAR in 1994 or Regulation Fair Disclosure (Reg FD) in 2000], ChatGPT likely helps retail investors better interpret disclosures that were

previously difficult to understand. This raises an important question in financial economics: can advances in LLMs reduce information asymmetries that stem not from unequal access to information but from unequal information-processing abilities?

To address this question and related issues, we develop the Coverage Score (*CS*) and its average value for each firm, *CSA*, which captures how much an otherwise uninformed trader can learn about a firm's financial health and performance from public disclosures using ChatGPT. We obtain the score by standardizing Compustat financial statement data and inputting it into OpenAI's GPT-3.5 Turbo model using a structured prompt designed to assess how effectively the model can help a novice investor interpret the information. In each of the eight runs at four different temperature levels (0.0 to 1.0), the model generates a *CS* from 0 to 100 based on: (i) the breadth of the data, meaning how comprehensive and detailed the disclosures are; and (ii) the usefulness of the data for decision-making, indicating how effectively ChatGPT can interpret the data for a novice. To obtain the *CSA* for each firm with minimal variation, we average the eight scores. Firms with high *CSA* provide detailed, nuanced disclosures that enable an LLM to extract, synthesize, and clarify economically meaningful information more effectively than simple rule-based methods. These disclosures add depth and detail, giving the LLM more opportunities to add value. Conversely, firms with low *CSA* provide sparse or very straightforward disclosures, leaving little room for the LLM to generate additional insights. Differences in *CSA* across firms help us identify treatment firms versus control firms in this study.

After identifying treatment firms based on *CSA*, we conduct DiD analyses to evaluate the impact of ChatGPT's launch on 10 outcome variables by comparing changes in these variables between treatment and control firms during the post-ChatGPT period. The 10 outcome variables include the posterior probability of informed trading computed following Brennan, Huh, and Subrahmanyam (2018), the price-impact parameters estimated based on Glosten and Harris (1988) and Sadka (2006), the high-frequency volatility measures computed using transaction prices and quote midpoints, and the absolute order imbalance as well as its two components (imbalances stemming from retail investors and non-retail investors). We estimate these variables daily.

Our results reveal a consistent pattern of information democratization following the release of ChatGPT. First, the treatment effect on informed trading is negative and statistically significant. For example, treatment firms experience a 5.2% reduction in informed trading. We also find that this decline mainly results from trades based on good news: the probability of informed buying decreases by about 7.9%. In contrast, the probability of informed selling remains mostly unchanged. These findings highlight the different effects of ChatGPT on informed trading. When faced with positive signals, retail investors engage in a deliberative process that involves evaluating fundamentals, valuation metrics, and qualitative disclosures. With ChatGPT's help, retail investors can narrow the gap in their disadvantage relative to

analysts and institutional investors. Conversely, when facing potential losses, this deliberation process may break down, leading them to skip analytical tools like ChatGPT and act on instinct. As a result, the selling advantage of sophisticated traders is likely to stay largely intact. Overall, analyzing the good- and bad-news components of informed trading shows that ChatGPT's effects function through behavioral selectivity. AI tools can level the playing field when cognitive skills are the main limitation, but have only a modest impact on the psychological barriers that cause loss-related biases.

We find a significant decrease in the price impact of trades in the treatment group stocks following the release of ChatGPT. Together with the above results, the evidence points to a coherent mechanism through which ChatGPT operates: by improving investors' ability to interpret public disclosures, the model reduces informational advantages previously held by better-informed traders, dampens informed trading, and lowers execution costs for uninformed market participants. We also find statistically and economically significant decreases in short-horizon return volatilities for treatment firms following the introduction of ChatGPT. This result is consistent with improved dissemination and interpretation of public information, which alleviates overreaction and reduces frictions in price formation. Hence, ChatGPT contributes to more stable short-term price dynamics by helping market participants process complex disclosures more effectively.

Improved interpretation of public information also has implications for trading behavior. If AI tools help investors reach more similar conclusions, they may reduce noise trading and disagreement among market participants. Consistent with this idea, we find evidence of greater alignment in trading decisions following the introduction of ChatGPT. In particular, imbalances in buy- and sell-side activity increase for affected firms, indicating more coordinated responses to public information rather than fragmented or idiosyncratic trading. This pattern suggests that ChatGPT fosters convergence in beliefs by clarifying the economic implications of disclosures, leading investors to trade in the same direction.

We further show that this convergence is driven primarily by retail investors. When trading activity is separated by investor type, the post-ChatGPT increase in coordinated trading is substantially stronger among retail traders than among non-retail traders. While both groups benefit from improved information processing, the effect is markedly larger for retail investors, who are more likely to face constraints in analyzing complex financial statements on their own. This evidence indicates that ChatGPT disproportionately enhances the ability of less-sophisticated investors to interpret public information, thereby narrowing interpretation gaps among market participants and improving price discovery.

We perform a series of robustness checks to ensure that spurious trends or modeling choices do not influence our results. Placebo tests using alternative event dates show no comparable effects before ChatGPT's introduction, supporting a causal interpretation. Additionally, our findings remain consistent when we use an alternative method to identify affected firms based on a smoothed exposure measure.

Throughout these checks, both the qualitative and quantitative conclusions stay the same, reinforcing that ChatGPT has significantly changed trading behavior by enhancing investors' understanding of public disclosures.

Our research indicates that generative AI democratizes information processing by creating a new group of "AI-assisted informed" participants. However, AI tools also influence firm fundamentals through changes in production processes and labor structures. Eisfeldt et al. (2025) quantify this impact using a metric called "labor (workforce) exposure to generative AI" (*LExpo*) and find significant valuation effects for firms with high *LExpo*. This highlights another dimension of AI's influence on industries and capital markets. For most outcome variables, we observe that the *CSA*-based effects of ChatGPT are similar for firms with high and low labor exposure to generative AI, supporting the view that *CSA* captures an information-processing channel related to the interpretability of existing public disclosures. Nevertheless, the Sadka permanent price-impact measure suggests that when valuations become more anticipatory and transformation-driven, AI-related labor exposure can partially diminish the disclosure benefits linked to high *CSA*. In narrative-driven environments (Shiller, 2017), sophisticated investors hold an advantage, limiting how much ChatGPT can lower information asymmetry.

This paper makes several contributions to the literature. First, our *CSA* metric establishes a replicable framework for measuring how effectively AI can extract and explain a firm's public information, providing a methodological foundation for research on corporate disclosures. We use iterative refinement and multi-temperature sampling to ensure consistent scores across different runs. Second, our study advances market microstructure research by identifying a new channel through which informational efficiency improves: the release of ChatGPT reduces information asymmetry, positioning AI tools as a key factor influencing market quality. Third, it contributes to the retail trading and behavioral finance literature by demonstrating that AI alleviates individual investors' cognitive limitations. While previous work (Barber and Odean, 2000, 2008) attributes retail underperformance to attention constraints and weak analytical skills, our evidence suggests that LLMs help mitigate these frictions and enable retail investors to become better informed. This paper also contributes to research on retail investor behavior in contexts where AI-driven coordination can help stabilize financial markets.

Additionally, unlike prior work (Lopez-Lira and Tang, 2023), our study advances the emerging LLM literature by shifting the focus from prediction to interpretability. We emphasize AI tools' explanatory power by using ChatGPT not as a black-box predictor but as an analytical assistant that helps clarify public information. Ultimately, our findings inform policy discussions on market fairness and disclosure practices. If AI enables retail investors to better understand complex filings, regulators might consider revisiting the "plain-English" disclosure requirements and encouraging more comprehensive, machine-readable text that benefits both human and AI users, thereby setting a new standard for "AI-compatible" disclosures.

Within the taxonomy of Eisfeldt and Schubert (2024), our paper falls into the category of generative AI used as a research tool for measurement rather than prediction. We use ChatGPT to construct a firm-level index that captures the interpretability of public disclosures for non-expert investors, extending prior applications that use LLMs to score exposure or task content.

The rest of the paper is organized as follows: Section 2 explains the methodology used to derive the *CSA* metric. In Section 3, we develop hypotheses linking ChatGPT's release to information asymmetry, informed trading, price impact, return volatility, and order imbalances. Section 4 describes the data and how variables are constructed. Section 5 presents the main findings. Section 6 tests the robustness of our results. Section 7 examines the combined effects of *CSA* and labor exposure to generative AI. Section 8 provides the conclusion.

2. Methodology for Capturing AI's Impact: Coverage Score Average (*CSA*)

In this section, we introduce the paper's central concept — the Coverage Score (*CS*) and its average value for each firm (*CSA*) — which measures where ChatGPT can most effectively assist uninformed investors. We use this concept to develop our testable hypotheses in Section 3. This metric is motivated by the ongoing information-processing gap between sophisticated (informed) and naive (uninformed) market participants. In traditional microstructure theory, informed traders gain from superior information, while uninformed traders incur adverse-selection costs that widen bid-ask spreads and reduce liquidity (Kyle, 1985; Glosten and Milgrom, 1985). Retail investors generally fall into the latter group and often underperform in trading because they lack the expertise or training to interpret complex financial reports (Barber and Odean, 2000). The Coverage Score aims to directly measure this gap by identifying firms with rich but complex public disclosures — cases where the informational divide between professional and retail investors is widest and where an AI assistant could most effectively narrow it. If ChatGPT truly helps level the playing field for firms with high coverage scores, we would expect to see significant changes in measures related to information asymmetry, return volatility, and order imbalances, as previously uninformed investors become better informed.

CSA measures how well a company's financial statements can be understood with the help of an LLM like ChatGPT. It essentially assesses how much useful insight an AI assistant can extract from a firm's disclosures for a novice investor. The score highlights companies where AI-driven analysis provides the most significant benefit to uninformed traders. Although any investor can gain insights from an AI-generated summary, those with limited financial knowledge benefit the most, as they start from a lower level of understanding. This idea aligns with evidence that generative AI tools disproportionately improve the performance of less-skilled users, reducing the gap between novices and experts (Noy and Zhang, 2023).

An important feature of the Coverage Score is that it is intentionally conditional on the interpretive capabilities of ChatGPT, the generative AI tool whose public release constitutes the information shock studied in this paper. Our objective is not to measure an objective or intrinsic notion of disclosure quality, but rather to capture how much economically meaningful information an AI assistant can extract and explain from a firm's public disclosures for an uninformed investor. In this sense, the Coverage Score reflects AI-amenable interpretability rather than disclosure quality per se. This distinction is central to our analysis: if generative AI lowers the cost of interpreting complex public information, then firms whose disclosures are especially conducive to AI-assisted explanation should experience the largest changes in information asymmetry and trading behavior following the public release of ChatGPT. Measures of information environments in finance — including analyst coverage, readability indices, and narrative complexity — are similarly conditional on the information-processing technologies available to investors. The Coverage Score differs only in that it explicitly conditions on the capabilities of a generative AI model, which is the technology shock of interest in this study.

To illustrate, consider two firms: one with minimal and straightforward financial statements, and the other with extensive and complex statements. A naïve investor reading on their own might only understand basic figures from either report. However, with ChatGPT's help, the investor could gain much more insight from the complex report than from the simple one. In other words, ChatGPT offers the most significant added value when a firm's disclosures are large but not easily understandable by non-experts. Conversely, if a firm's public information is very limited or straightforward, an LLM adds little because there is not much hidden insight to uncover. Therefore, we can use *CSA* to differentiate the treatment group from the control group of firms. High-*CSA* firms are those for which ChatGPT's assistance is expected to reduce information asymmetry, while low-*CSA* firms serve as a control group with limited AI impact.

A key advantage of the *CSA* metric is that it avoids ad hoc classifications based on firm size or industry. Because an LLM evaluates each firm's disclosure complexity directly, the group of high-score firms is diverse: it includes both large- and small-cap companies across many industries, rather than concentrating on any one sector or industry. This suggests the score captures a broad notion of disclosure richness and LLM amenability — how conducive a firm's information environment is to AI interpretation — rather than proxying for a single firm characteristic. The cross-cutting nature of the metric enhances its validity: differences in market outcomes between high- and low-score firms can be more confidently attributed to AI-facilitated understanding rather than to confounding factors such as firm size or visibility.

Additionally, our methodology uses AI to measure information interpretability. Instead of viewing ChatGPT as a crystal ball predictor, we treat it as an analytical assistant. By prompting the model to summarize each firm's financial statements, we derive a score based on the insight it provides to a novice reader. This method provides a clear and repeatable measure of the amount of "explainable" information

contained in each firm's disclosures. In summary, *CSA* is a straightforward indicator of where generative AI has the highest potential to democratize information. It forms the basis for our empirical tests, enabling us to investigate whether firms with a high score experience more significant changes in information asymmetry, informed trading, and related outcomes after ChatGPT's public release.

In our empirical tests, we classify firms into two groups based on *CSA*. Firms with higher *CSA* — those with information-rich but complex disclosures that are especially suitable for AI-assisted interpretation — constitute the treatment group. In comparison, firms with lower scores form the control group. Importantly, the Coverage Score is built solely from information available before ChatGPT's public release and is therefore fixed relative to the outcome variables we analyze.[1] We provide a detailed explanation of the Coverage Score, including prompt design, scoring criteria, and the exact cutoff between treatment and control, in Section 4 and Internet Appendix A. This setup allows us to develop testable hypotheses about how generative AI influences information asymmetry and trading behavior, before focusing on the technical implementation.

3. Hypothesis Development

Our primary objective in this study is to examine whether the release of ChatGPT has affected information asymmetry, return volatility, and order imbalances in firms with high *CSA*. For this purpose, we first develop a set of testable hypotheses linking the availability of generative AI to changes in key measures of informed trading activity.

3.1. Information Asymmetry and Informed Trading

3.1.1. Informed Trading

Information asymmetry, formalized by Akerlof (1970) and operationalized in market microstructure models by Kyle (1985) and Glosten and Milgrom (1985), arises when market participants possess unequal access to value-relevant information. Prior research distinguishes between asymmetry in private information stemming from proprietary or insider signals and asymmetry in public information, which occurs when data are publicly available but investors differ in their skills in processing and interpreting it. While most empirical studies have focused on private information asymmetry, other studies highlight the importance of disparities in public information processing, particularly among retail investors.[2]

[1] See OpenAI documentation for GPT-3.5 Turbo's knowledge cutoff date (September 1, 2021): https://platform.openai.com/docs/models/gpt-3.5-turbo?utm_source=chatgpt.com

[2] See, for example, Hong and Stein (2007) and Tetlock (2007).

Even when disclosures are publicly accessible (e.g., under Reg FD), differences in cognitive and processing capacity can still limit the ability of unsophisticated traders. Sophisticated investors rely on analytical teams, proprietary valuation models, and advanced computational systems to extract insights from financial disclosures. In contrast, retail investors face resource and comprehension limitations that hinder effective information processing (Hirshleifer and Teoh, 2003). This gap raises adverse-selection costs, discouraging retail participation and impairing price discovery.

ChatGPT's release marks a significant shift in this balance. Unlike earlier technologies that improved access to information (e.g., EDGAR in 1994 and algorithmic newsfeeds in the 2000s), LLMs have greatly enhanced comprehension of information. ChatGPT allows users to summarize lengthy 10-K reports, translate technical accounting language into simple English, benchmark performance, and synthesize complex narratives — tasks that once required specialized expertise. As a result, these capabilities may now be democratized among market participants, since AI tools are available to anyone with an Internet connection.

Retail investor adoption of ChatGPT suggests two equilibrium adjustments. First, AI tools flatten the learning curve, allowing uninformed traders to interpret disclosures more accurately. Second, the comparative advantage of informed traders erodes, potentially reducing their market participation. Consequently, we conjecture that information asymmetry and informed trading decline in financial markets at least in the short run, given ChatGPT's rapid adoption by 100 million users within two months (Reuters, 2023). These effects should be most pronounced for firms to which ChatGPT adds the greatest informational value, namely those with high *CSA*. These considerations lead to our first hypothesis:

Hypothesis 1: *Following ChatGPT's release, the probability of informed trading decreases more for treatment firms (with high CSA) than for control firms (with low CSA).*

A natural prior is that improved information technologies should increase information asymmetry and informed trading. This view, however, conflates the level of information in the market with the dispersion of information across traders. In standard microstructure models, information asymmetry arises from heterogeneity in investors' information or interpretation abilities, not from the absolute amount of information available. Before ChatGPT, superior information processing was concentrated among a small subset of sophisticated investors, generating persistent informational advantages. ChatGPT primarily operates on this relative dimension by lowering the cost of interpreting complex public disclosures for less-sophisticated traders. As a result, posterior beliefs conditional on public information become more homogeneous across market participants, reducing the scope for profitable informed trading. In equilibrium,

when informational advantages erode because many traders interpret the same public information similarly, measures of informed trading decline (although overall information processing in the market improves).

3.1.2. Asymmetry in the Effects: Informed Trading on Good News vs. Bad News

Studies document that information processing is not symmetric across positive and negative events. Behavioral studies grounded in prospect theory and cognitive load theory suggest that retail investors deliberate more when evaluating good news than bad news.[3] Positive information, which signals potential gains, induces analytical evaluation and confirmation seeking (Barber and Odean, 2008). Conversely, negative information often elicits immediate reactions driven by fear of loss, prompting rapid liquidation ("sell first, analyze later") and reducing reliance on analytical tools (Shefrin and Statman, 1985). The disposition effect further magnifies this asymmetry: investors tend to hold losing positions too long and sell winners too early (Shefrin and Statman, 1985; Kaustia, 2010). That is, AI's role in investment decisions varies across contexts. When investors face unrealized losses, cognitive dissonance and regret avoidance often suppress analytical engagement. Conversely, during positive news events, motivated reasoning and confirmatory analysis tend to increase reliance on AI tools, as investors use them to validate and reinforce optimistic interpretations.

Another relevant factor reinforcing the disparity or asymmetry is short-sale constraints, which make it difficult for traders to take advantage of bad news. Therefore, stocks that are hard to short tend to be overvalued and exhibit lower subsequent returns (Jones and Lamont, 2002). As AI tools induce the democratization of information, we expect that the unwinding or offsetting effects should be stronger when trading on good news. All the above aspects suggest that ChatGPT's democratizing effects should be greater for informed trading on good news than on bad news. Thus, we test the following hypothesis:

Hypothesis 2: *The decline in the probability of informed trading after ChatGPT's introduction is more pronounced for good-news events than for bad-news events.*

3.1.3. Price Impact

Building upon the expectation that ChatGPT's release may reduce informed trading, we examine whether lower information asymmetry translates into reduced adverse-selection costs and market frictions. Market microstructure theory posits a direct link between informed trading and price impact. The Glosten-Harris (1988) model captures this relationship through the price-impact parameter (λ^{GH}), while Sadka (2006) extends the framework by decomposing trading costs into four more granular components,

[3] Behavioral finance indicates that investors process positive and negative information asymmetrically, influencing trading patterns and price discovery (Kahneman and Tversky, 1979; Hirshleifer, 2001).

compared to two in Glosten and Harris (1988). If ChatGPT reduces informed trading among treatment firms, market makers should perceive lower risk and respond by narrowing bid-ask spreads. Consequently, the adverse-selection component of trading costs measured by λ^{GH} or λ^{S} should decline in the post-ChatGPT period. Accordingly, we propose the following hypothesis:

Hypothesis 3: *Following ChatGPT's release, the adverse-selection component of trading costs decreases in the treatment group of firms relative to the control group of firms.*

3.2. Return Volatility

Retail investors often contribute more to volatility through impulsive or uninformed trades. As ChatGPT became available to retail investors, it provides them with easier access to information and analysis tools. This narrows the gap between informed and uninformed traders, reducing uncertainty in price discovery. With AI assistance, decisions may also become more consistent and data-driven, reducing erratic price movements. Furthermore, when more participants have access to similar analytical capabilities, prices incorporate information more quickly and smoothly. This reduces large price swings caused by misinterpretations. These considerations lead to the following hypothesis:

Hypothesis 4: *Following ChatGPT's release, return volatility declines in the treatment group of firms relative to the control group of firms.*

3.3. Coordination in Trading and Order Imbalances

Prior to the advent of ChatGPT, investors frequently maintained divergent opinions and employed opposing trading strategies despite having access to identical information. ChatGPT has the potential to diminish this diversity of opinion through two primary mechanisms. Firstly, by serving as an understanding channel: AI tools such as ChatGPT assist investors in comprehending complex financial disclosures—translating technical accounting terminology, summarizing risk factors in 10-K filings, and identifying key performance indicators ― thereby bridging comprehension gaps that previously differentiated retail investor responses. By elucidating tone, context, and the implications of management commentary, ChatGPT enables investors to develop more accurate and coordinated perceptions of a company's fundamentals.

Second, the convergence channel: because ChatGPT relies on a consistent LLM architecture and shared training data, users analyzing the same information tend to have similar interpretations and reasoning paths. This homogenization of analytical framing reduces idiosyncratic variations in investor judgments. As a result, investors are more likely to reach the same directional conclusion, positive or negative, when

evaluating identical disclosures. Together, these channels decrease interpretive heterogeneity, foster collective understanding, and improve coordinated trading around public information disclosures. For example, if investors previously split evenly (50/50) in response to earnings news but now converge toward an 80/20 consensus post-ChatGPT, the absolute order imbalances should grow, indicating better coordination and faster information incorporation. Based on these considerations, we propose the following hypothesis:

Hypothesis 5: *Following ChatGPT's release, the absolute order imbalance increases in the treatment group of firms relative to the control group of firms.*

The above prediction might seem counterintuitive, given the anticipated declines in informed trading after ChatGPT's introduction. However, these effects can occur together because of compositional shifts: traditional informed traders' activities decrease, while coordinated trading by retail investors rises. The overall result depends on whether retail coordination exceeds the decline in informed order flows. If ChatGPT effectively enhances retail understanding, the total imbalance should grow following its launch.

To investigate this issue more thoroughly, we classify trades into retail and non-retail categories following Barber et al. (2024), who improve upon the method developed by Boehmer et al. (2021), and calculate the absolute order imbalance for each category. This method enables us to determine whether the coordination is mainly driven by retail investors. Consequently, we test the following hypothesis:

Hypothesis 6: *Following ChatGPT's release, the absolute order imbalance originating from retail traders increases in the treatment group of firms, relative to the control group of firms. Moreover, the degree of increase is stronger for the order imbalance stemming from retail traders than from non-retail traders.*

4. Data and Construction of Variables

Our empirical analyses use data on U.S. common stocks listed on the NYSE, AMEX, and Nasdaq, covering the two years surrounding the public release of ChatGPT on November 30, 2022. We interpret this event as a distinct accessibility and salience shock that made large language model capabilities widely visible and usable by retail investors. The pre-event window spans from November 30, 2021 to November 29, 2022, while the post-event window extends from December 1, 2022 to November 29, 2023. Below, we detail how we construct the coverage score, measures related to informed trading, price impact, order imbalance, return volatility, and other firm characteristics. The primary outcome variables are estimated or

computed daily at the stock level using the NYSE Trades and Automated Quotations (TAQ) database, along with data from Compustat and CRSP.

4.1. Coverage Scores

4.1.1. Calculation of Coverage Scores

We calculate the Coverage Score (*CS*) for each firm by feeding its financial statements to an LLM (ChatGPT) under a standardized prompt and obtaining the AI's judgment on two aspects: (1) the *coverage* of data contained in a firm's financial statements, that is, how complete and comprehensive the available financial information is; and (2) the *usability* for decision-making, that is, how helpful the information available in the data is, so that the AI can interpret and advise the AI user about the direction of trading. We use the Compustat database to extract each firm's annual SEC 10-K filing for 2022, which contains balance sheets for 2021–2022 and income statements for 2020–2022. Each firm's statements are then mapped into a standardized list of financial items, enabling the LLM to observe both reported and missing elements. Missing items are explicitly marked as such rather than omitted, which allows the model to assess not only what is disclosed but also what is absent, thereby capturing the dimension of completeness in its financial reporting.[4] This stage produces a uniform schema that serves as the model's input template. The model also generates a numeric score (*CS*) from 0 to 100 based on the criteria stated in the prompt. A firm with a *CS* of 100 indicates that its financial statements are complete in coverage and usability, enabling ChatGPT to advise uninformed investors. On the other hand, a firm with a *CS* of 0 represents the opposite extreme, posing significant challenges for ChatGPT to interpret or assist the AI user.

To be specific, we design the prompt (input text or instructions for the OpenAI API) so that it can generate scores of 50 points each for data coverage (how complete the data is) and data usability for decision-making (how helpful ChatGPT could be). The score is evenly divided between the breadth and depth of available information and the practical usefulness of that information when interpreted by ChatGPT. While the Coverage Score is based on these two conceptual dimensions, we use their equally weighted average as our main measure because our goal is to reflect overall exposure to AI-assisted interpretation rather than to evaluate the dimensions separately. Since no standard exists for assessing firm-level characteristics using large language models, we adopt a structured prompting approach that guides the model step by step through the evaluation process before requesting a final numeric score. This method produces highly consistent results: given the same inputs and criteria, the model yields scores in a consistent format and applies the evaluation dimensions reliably across firms. We manually verified this process by

[4] Feldman et al. (2023) show that supplying context with embedded tags significantly reduces hallucination rates by helping LLMs recognize when information is missing or when they are operating outside their domain knowledge.

reviewing a subset of the model's outputs and corresponding firm disclosures and confirmed that the generated scores match the intended criteria.

Additionally, to ensure statistical reliability and reduce LLM randomness, we use a multi-temperature approach: eight separate runs, or two runs at each of four different temperatures (0.0, 0.5, 0.7, and 1.0) for each firm. The average (i.e., *CSA*) of the eight scores (*CS1*-*CS8* reported in Table 1) serves as our main metric for grouping the treatment and control firms. We find that averaging these scores significantly decreases variability, making the *CSA* metric more reliable. In calculating *CS*s, we use OpenAI's GPT-3.5 Turbo as our LLM for two reasons: it was the most accessible model to retail investors during the post-event window, and its training cutoff predates the release of the 2022 Form 10-K filings used in our analysis, eliminating the risk of post-event disclosure contamination.

Because the GPT API does not allow sampling to be fully disabled, individual model outputs are inherently random. While purely deterministic settings might be suitable for some measurement tasks, our goal is to reflect how investors plausibly use generative AI in practice. Users interact with LLMs using diverse prompts and parameter settings, from highly deterministic summarization to more exploratory analysis. Evaluating firms across multiple temperature settings helps us capture this variation while ensuring robustness through averaging.

Internet Appendix A offers detailed technical information, including the complete data-preprocessing protocol, prompt wording, system parameters, example model outputs, and robustness checks. It also includes descriptive statistics and visual examples showing how *CS* and *CSA* differ across firms and model setups. These additional materials ensure full transparency of the methodology while allowing the main text to focus on the conceptual framework and empirical analysis.

4.1.2. Grouping the Treatment vs. Control Firms

The next step is to identify the treatment group of firms from the others to properly assess the impact of ChatGPT. A straightforward method would be to divide the sample at the median *CSA*. However, this approach has two issues. First, it creates a discontinuity at an arbitrary point in the distribution, which can misclassify firms near the cutoff whose treatment effects differ only slightly. If the treatment effect changes smoothly with *CSA*, a median split ignores meaningful variation, reducing statistical accuracy (Imbens and Lemieux, 2008). Second, as shown in Table 1 and Figure A5 in Internet Appendix A, the distributions of the eight coverage scores (*CS1*-*CS8*) are generally bimodal and left-skewed, with a large probability around the median. Therefore, assigning treatment based on the median could significantly bias our results (Angrist and Pischke, 2009, Chapter 4).

A more reasonable approach is to identify treatment firms using upper-percentile cutoffs, such as the top quartile or quintile, and limit the treatment group to firms with clearly high exposure. This method

offers several benefits. For example, it enhances the validity of grouping by focusing on firms where ChatGPT's impact is most significant, such as those with complex information environments, high retail investor interest, and substantial textual disclosures that AI can process efficiently. Based on the points above (bimodality in Figure A5 and left-skewness in Table 1), we select the 70th percentile of *CSA* as our cutoff: the top 30% of firms form the treatment group, while the remaining 70% serve as the control group. Table 1 shows the average *CSA* is 52.13, with data available for more than 2,600 firms. Individual *CS* scores range from zero to 90, with the maximum score generally increasing with temperature. We find that the 70th percentile of the *CSA* distribution is 58.40, which aligns with the bin containing the highest proportion in Figure A5 (see the bottom-right histogram). Considering the eight individual distributions in Figure A5, we believe 58.40 is a natural cutoff point for defining treatment firms.

4.2. Measures of Informed Trading, Information Asymmetry, Volatility, and Order Imbalances

4.2.1. Daily Posterior Probabilities of Informed Trading

Our first proxies for informed trading are the daily posterior probabilities of informed trading on good news, on bad news, and their sum. Easley, Kiefer, O'Hara, and Paperman (EKOP) (1996) develop a model in which one of three events occurs on each day: no news ($\emptyset$), good news (g), or bad news (b). Then, we can express the unconditional probabilities of the above events as $Pr(\emptyset) = (1-\alpha)$, $Pr(g) = \alpha(1-\delta)$, and $Pr(b) = \alpha\delta$, respectively, where α is the probability that an information event occurs on the day, while δ is the probability that the event is bad news. On a given trading day, if an information event occurs, it is observed only by informed traders who trade on this information: if a bad-news (good-news) event occurs, the informed traders sell (buy) at the rate μ. Regardless of any news event, uninformed traders buy and sell at the rates ϵ_B and ϵ_S, respectively. EKOP show how to estimate the model parameters (α, δ, μ, ϵ_B, and ϵ_S) from a time series of the numbers of daily buyer- and seller-initiated trades. Using the parameters, EKOP compute the unconditional probability of informed trading (*PIN*).

In the literature, however, some issues with *PIN* have been raised (e.g., too low frequencies, no buy/sell direction, etc.). Building on the intuition of EKOP, Brennan, Huh, and Subrahmanyam (BHS, 2018) develop high-frequency (daily) measures of informed trading. Based on the methods of BHS, we construct the daily posterior probabilities of informed trading on good news (π_g) and bad news (π_b) separately, conditional on observing the daily numbers of buyer-initiated trades (B) and seller-initiated trades (S). The daily posterior probabilities of informed trading on good and bad news are defined as:

$$\pi^g \equiv Pr(g|B,S) = \frac{\alpha(1-\delta)(\epsilon_B+\mu)^B\epsilon_S^S}{\alpha(1-\delta)(\epsilon_B+\mu)^B\epsilon_S^S + \alpha\delta\epsilon_B^B(\epsilon_S+\mu)^S + (1-\alpha)e^{\mu}\epsilon_B^B\epsilon_S^S}, \quad (1)$$

$$\pi^b \equiv Pr(b|B,S) = \frac{\alpha\delta\epsilon_B^B(\epsilon_S+\mu)^S}{\alpha(1-\delta)(\epsilon_B+\mu)^B\epsilon_S^S + \alpha\delta\epsilon_B^B(\epsilon_S+\mu)^S + (1-\alpha)e^{\mu}\epsilon_B^B\epsilon_S^S}. \quad (2)$$

Given that intraday order flows (and daily aggregated buys and sells) are available, we first estimate the five parameters (α, δ, μ, ϵ_B, and ϵ_S) following Hwang, Lee, Lim, and Park (2013) on a monthly basis using a three-month rolling window. The monthly (rather than 'annual', as in other studies) estimation allows us to estimate parameters that capture the time-varying features of information events and trading activities. Using the monthly estimates of the five parameters from the previous month, we can compute the daily posterior probabilities of informed buying (π_g) and selling (π_b) as in Eqs. (1) and (2) each day using the numbers of daily buys (B) and sells (S) in the current month. We then obtain the probability of informed trading, π^e, as the sum of π^g and π^g to capture daily (total) informed trading activities.

4.2.2. Price-Impact Measures Based on Glosten and Harris (1988) and Sadka (2006)

We use two types of price-impact measures to capture information asymmetry or the adverse-selection component of trading costs. Based on Glosten and Harris (1988), we first decompose the total trading costs into the following two components:

$$\Delta P_{i,k,d} = \bar{\varphi}_{i,d}^{GH}(S_{i,k,d} - S_{i,k-1,d}) + \lambda_{i,d}^{GH}(S_{i,k,d}V_{i,k,d}) + \xi_{i,k,d}, \quad (3)$$

where $S_{i,k,d}$ is equal to $+1$ if trade k is buyer-initiated on day d for stock i, and -1 if seller-initiated; $V_{i,k,d}$ is volume at trade k; $(S_{i,k,d}V_{i,k,d})$ is intraday order flows; $\bar{\varphi}_{i,d}^{GH}$ is the noninformation component of trading costs (i.e., order-processing and inventory-holding costs) for stock i on day d; and $\lambda_{i,d}^{GH}$ is the adverse-selection component.

Following Sadka (2006), we also estimate the unexpected order flows (τ_t), their variance (σ_τ^2), and the fitted value of order flows ($\widehat{S_tV_t}$) after filtering raw order flows by the AR(5) process. The unexpected sign (ω_t) of a trade is computed using the following equation:

$$\omega_t = S_t - E_{t-1}(S_t) = S_t - \left\{1 - 2\Phi\left(-\frac{\widehat{S_tV_t}}{\tau_t}\right)\right\}, \quad (4)$$

where $\Phi(.)$ denotes the normal cumulative distribution function. Then, we decompose trading costs into four elements by running the following time-series regression:

$$\Delta P_{i,k,d} = \bar{\lambda}_{i,d}^{S}(S_{i,k,d}V_{i,k,d} - S_{i,k-1,d}V_{i,k-1,d}) + \lambda_{i,d}^{S}\tau_{i,k,d} + \varphi_{i,d}^{S}\omega_{i,k,d} + \bar{\varphi}_{i,d}^{S}(S_{i,k,d} - S_{i,k-1,d}) + \xi_{i,k,d}, \quad (5)$$

where $\bar{\lambda}_{i,d}^{S}$ is the transitory variable cost for stock *i* on day *d*; $\lambda_{i,d}^{S}$ denotes the permanent variable cost; $\varphi_{i,d}^{S}$ is the permanent fixed cost; and $\bar{\varphi}_{i,d}^{S}$ denotes the transitory fixed cost.

To estimate Glosten-Harris lambda (λ^{GH}) and Sadka lambda (λ^{S}), as well as other components, we run the time-series regressions as in Eqs. (3) and (5) (with a constant term) using all the intraday order flows and related variables available within day *d*. Note that the Sadka model decomposes trading costs into four components. Since we are interested in the trading cost related to information asymmetry, we use λ^{GH} and λ^{S} that represent the permanent variable cost, screening out other components of trading costs. Another issue that arises when estimating the measures in Eqs. (3) and (5) is whether to use dollar order flows or share order flows. We use dollar order flows for estimation, as we believe they provide a better estimate of the price-impact measures.

4.2.3. Return Volatility

Studies document that more efficient markets exhibit lower return variance in general, as prices quickly incorporate information. Since volatility captures periods of heavy information flow or uncertainty, high-frequency return volatility has often been used as a proxy or diagnostic for market efficiency (e.g., Chordia, Roll, and Subrahmanyam, 2005). To gauge the impact of generative AI on price movements and market efficiency, we compute two types of high-frequency return volatility. The first is the daily standard deviation of five-minute returns based on transaction (observable) prices (denoted as $Vola_{5m}$). This is to capture the variability in traded prices at a high frequency. In order to reduce the potential influence of bid-ask bounce, we compute the daily standard deviation of five-minute returns based on bid/ask-quote midpoints (denoted as $Vola_{5m}^{mid}$) as well.

To estimate the daily posterior probabilities, price-impact measures, high-frequency volatility measures, and absolute order imbalances mentioned above, we need to process intraday order flows by matching each trade with relevant ask/bid quotes for each stock listed on the NYSE, AMEX, or Nasdaq during the sample period (11/2021-11/2023) using the TAQ database. To obtain raw order flows, intraday trades and quotes from TAQ are matched, and each trade is classified as buyer- or seller-initiated. For this purpose, we use the Holden and Jacobsen (2014) algorithm [with a nanosecond-delay rule using the 'daily' TAQ database (DTAQ)], which employs time interpolation and matches trades with national best bid and offer (NBBO) quotes across all exchanges. After matching trades and quotes, if a trade occurs above (below) the quote midpoint, it is considered buyer-initiated (seller-initiated).

4.2.4. Absolute Order Imbalances

Order imbalances are often used to explain returns or trading behavior (Chordia, Roll, and Subrahmanyam, 2002; Chordia and Subrahmanyam, 2004). Using available intraday order flows, we aggregate buy and sell trades within each day and calculate the daily absolute order imbalance, defined as $ADOI = \frac{|dbuy-dsell|}{dbuy+dsell}$, where *dbuy* and *dsell* are the daily dollar values of shares bought and that of sold, respectively. Democratization of information is expected to reduce information asymmetry and informed trading among market participants. On the other hand, the availability of AI may cause their view about a firm's future cash flow and, hence, their reactions to such a prediction to be more homogeneous. We use order-imbalance measures to capture this effect by *ADOI*.

To examine the issue, we calculate the measures separately by trader type: the absolute order imbalances caused by retail investors (denoted $ADOI^R$) versus non-retail investors ($ADOI^{NR}$). This helps us differentiate between trading activities from different market participants, since institutional and retail traders likely have different information sets and exhibit distinct trading behaviors. To identify retail versus non-retail trades in the TAQ database, we use the algorithm developed by Barber et al. (2024), which improves upon the method proposed by Boehmer et al. (2021).

4.2.5. Firm Characteristics

In addition to the measures mentioned above, we calculate other variables known to explain or predict information asymmetry, trading activities, and return volatility. As one of the control variables, we measure firm size as the natural log of daily market capitalization ($SIZE$) where market capitalization equals the daily closing price multiplied by the number of shares outstanding, obtained from CRSP. As an additional control variable, we use the book-to-market ratio (BTM) reported in the Compustat quarterly database. To avoid look-ahead bias while maintaining daily frequency, we match the previous quarter-end book value with daily market-equity values within each quarter, ensuring that each trading day a firm has the most recent book-to-market ratio.

Wherever relevant, we also control for volatility, which is calculated from high-frequency transaction data. Specifically, we use the daily standard deviation of intraday returns (denoted as $VOLA$) computed by WRDS (for details, see Formula 23 in its user manual titled, "WRDS Intraday Indicator Data"). By controlling for the above firm characteristics, we aim to isolate the effect of AI on variations in informed trading, price impact, return volatility, and order imbalances during the post-ChatGPT period.

4.3. Descriptive Statistics

Except for the total number of observations (Obs) and the average number of firms per month (NFirms/mon), each of the five statistics (mean, median, STD, skewness, and kurtosis) is first calculated

across all sample firms for each day. We then report the time-series average of these daily statistics in Table 2. The total number of observations (Obs) used in this study ranges from 0.784 to 1.211 million firm-days.

Panel A shows that the average daily posterior probability of informed trading (π^e) is approximately 32.1%. When broken down into the probabilities of informed buying (π^g) and selling (π^b), the latter is slightly higher than the former (15.7% vs. 16.4%). Due to well-known overflow and underflow problems, the average number of firms (NFirms/mon) with these three probability measures is smaller—about 1,560 firms—compared to more than 2,400 firms for other variables. Regarding the price-impact measures, it is known that the (raw) estimates of λ^{GH} and λ^S are highly leptokurtic (see Huh, 2014). Therefore, they are winsorized at the 0.5th and 99.5th percentiles. Because of their small values, they are also multiplied by 10^4. After winsorization, both measures become much less leptokurtic. The Sadka lambda (λ^S) is generally smaller than the Glosten-Harris lambda (λ^{GH}) (0.047 vs. 0.095), because the former is estimated by decomposing trading costs into four more granular components.

Panel B shows that the daily standard deviation of five-minute returns ranges from 0.036 to 0.052, with the one based on quote midpoints ($Vola_{5m}^{mid}$) being smaller than the one based on transaction prices ($Vola_{5m}$). Both volatility measures tend to be right-skewed and leptokurtic. Panel C shows that the daily absolute order imbalance (*ADOI*) is 16.1% on average for a typical firm during the sample period. Notably, the absolute order imbalance caused by retail investors ($ADOI^R$: 27.2%) is much higher than that caused by non-retail investors ($ADOI^{NR}$: 15.5%) or the overall measure (*ADOI*).

Panel D shows that the average market capitalization (*MV*) of our sample firms is $11.9 billion. Because *MV* is highly skewed and leptokurtic, we use the natural log of *MV* as a firm size variable (*SIZE*) in regression analyses. The book-to-market ratio (*BTM*) is about 0.668. The daily return volatility measure obtained from WRDS (multiplied by 10^2) (*VOLA*) is 0.004 on average.

5. Empirical Analyses

5.1. Method of Estimation

Given the *CSA* metric to identify the treatment group of firms, we examine the effects of ChatGPT's release on trading activities in those firms in the post-event period. For this purpose, we employ difference-in-differences (DiD) regressions as follows:

$$Y_{i,t} = \alpha + \beta_1(POST_t \times TREAT_i) + \beta_2 POST_t + \beta_3 TREAT_i + \sum_{k=1}^{3} \theta_k X_{ik,t} + \gamma_i + \delta_t + \varepsilon_{i,t}, \quad (6)$$

where $Y_{i,t}$ is one of the outcome variables for firm *i* at day *t*; $POST_t$ is an indicator variable equal to one if the observation occurs after the event date (November 30, 2022), and zero otherwise; $TREAT_i$ is an

indicator variable equal to one for firms in the top 30% of the *CSA* distribution, and zero otherwise; $X_{i,t}$ includes three firm characteristics [i.e., firm size (*SIZE*), book-to-market equity (*BTM*), and return volatility (*VOLA*)] as control variables; γ_i is the firm fixed effects; and δ_t is the date fixed effects.

The coefficient (β_1) on the interaction term (*POST* × *TREAT*) in Eq. (6) captures the additional change in the outcome variables for treatment firms, relative to that for control firms, following ChatGPT's release. We estimate two specifications: one includes date fixed effects to account for market-wide shocks and time trends, and another includes both firm and date fixed effects to control for time-invariant firm characteristics and market-wide shocks. The specification with both firm and date fixed effects offers the most rigorous estimation, isolating within-firm variations around ChatGPT's release while removing aggregate time trends. Standard errors are clustered by firm and date to address serial correlation within firms and cross-sectional dependence across dates. Because some of our dependent variables (e.g., π^e and *ADOI*) are bounded in [0, 1], we employ fractional-logit estimation, reporting the marginal effects of the variables in the interaction term (but not for other variables to save space). We use OLS when other measures are used as the dependent variable.

5.2. Effects of ChatGPT's Release on Informed Trading

5.2.1. Overall Effect

We first examine the impact of ChatGPT's release on informed trading without considering whether it is triggered by good news or bad news. The first two columns in Table 3 show the results for the probability of informed trading (π^e) when date fixed effects are included and when both firm and date fixed effects are included, respectively. In Column (1), the coefficient on (*POST* × *TREAT)* is −0.0877 with a *t*-statistic of −2.70 (significant at the 1% level), indicating that after November 30, 2022, informed-trading probability declines more for firms in the top 30% of the *CSA* distribution than for other firms. In Column (2), which adds firm fixed effects, the coefficient on (*POST* × *TREAT)* remains negative and significant (−0.0841, $t = -2.47$, significant at the 5% level). The near-identical magnitude across Columns (1) and (2) implies that the post-ChatGPT decline in informed trading among treated firms is not driven by time-invariant cross-sectional differences between high-*CSA* and low-*CSA* firms; instead, it reflects a within-firm change in informed-trading probability for treated firms relative to controls in the post-event period.

Panel B shows that the marginal effect of (*POST* × *TREAT)* on π^e is −0.0167 (significant at 1%) in Column (2). Interpreting this as an instantaneous change in the probability metric, the results imply that ChatGPT's release is associated with about a 1.67 percentage-point reduction in the estimated daily probability of informed trading for treated firms relative to controls. This is equivalent to a 5.2% decrease relative to the sample average of π^e (0.321) and a decline that constitutes 3.7% of its standard deviation (0.448) (see Table 2).

As to the control variables, firm size ($SIZE$) is positively associated with π^e, implying that larger firms are subject to more informed trading because they may attract greater investor attention despite higher levels of transparency in their disclosures. The effect of the book-to-market ratio (BTM) on π^e tends to be positive but rather weak. Return volatility ($VOLA$) plays no role.

The above result strongly supports Hypothesis 1: ChatGPT's launch helped reduce informed trading more significantly among firms with high *CSA*. The concentration of effects in disclosure-heavy firms suggests that ChatGPT's advantage is rooted in information processing rather than ease of access. If the main channel was improved accessibility — such as quicker retrieval of filings — we would see uniform declines across all firms. Instead, the decline is concentrated in firms with the most complex textual and qualitative statements, which aligns with LLMs' strengths in summarizing, contextualizing, and interpreting texts. This matches studies showing ChatGPT's skill in financial text analysis (Wu et al., 2023). By extracting sentiment from earnings-call transcripts, identifying risks in 10-K MD&A sections, and benchmarking performance across peers, ChatGPT can perform tasks once limited to professional analysts. As a result, this AI technology reduces the information-processing gap that once separated sophisticated investors from retail investors, effectively leveling the playing field for firms with the greatest comprehension challenges.

The above finding has implications for broader debates on artificial intelligence and financial democratization. Optimists argue that generative-AI tools may help narrow informational disparities between retail and institutional investors, whereas skeptics caution that they may exacerbate social uncertainty and reinforce concentrated technological power. Our evidence supports the optimistic view—at least in the short term—indicating that accessible AI tools significantly reduce information asymmetry and informed trading.

The implications also extend to corporate disclosure incentives. The above result shows that ChatGPT provides the most significant benefits to firms with detailed, information-rich filings. Historically, firms were encouraged to simplify their filings to accommodate retail investors' limited analytical skills, in line with "plain-English" initiatives (Bloomfield, 2002; Li, 2008). With AI-assisted understanding, retail investors can now grasp complexity rather than be discouraged by it. Firms aiming to share favorable information with market participants may have increased incentives to expand their disclosures, knowing that more detailed text improves interpretability through AI tools. These dynamics could create a positive feedback loop for transparency: as firms improve disclosure quality, AI-enabled investors respond with more accurate valuations and greater market efficiency.

5.2.2. Informed Trading on Good News versus Bad News

To examine whether ChatGPT's effects vary by news type (i.e., Hypothesis 2), we re-estimate the regression model using the probability of informed trading on good news (π^g) and bad news (π^b) separately. Columns (3) and (4) in Table 3 show the results for π^g. Panel A shows that the coefficients on the interaction term remain negative and statistically significant, and their sizes are similar between the two fixed effects specifications, indicating that time-invariant firm characteristics do not drive the post-ChatGPT decline in informed trading among treated firms. Panel B shows that the treated firms experience a drop of about 1.24 percentage points in their likelihood of informed buying on good news (π^g) relative to the control firms after ChatGPT's release. This translates to a 7.9% decrease relative to the sample mean of π^g (0.157) and amounts to 3.5% of its standard deviation (0.351). The decrease (7.9%) in π^g is larger than the decrease (5.2%) in π^ediscussed above. The results for the control variables are qualitatively identical to those when we use π^eas the dependent variable.

Columns (5) and (6) show that the treatment effect on π^b is weaker than the treatment effect on π^g shown in Columns (3) and (4). The regression specification with only the date fixed effects in Column (5) shows that the DiD coefficient is negative and significant at the 10% level. However, the other column indicates no significant relationship. This muted response supports our Hypothesis 2: when faced with negative signals, investors react differently, leaving ChatGPT's advantages unexploited.

The control variables also behave differently. The effect of *SIZE* is mixed, with its coefficient turning negative once the joint fixed effects are included, suggesting that informed selling becomes concentrated in smaller, less liquid stocks. BTM is negatively associated with π^b, consistent with the notion that value firms are less vulnerable to informed selling. In particular, the effect of *VOLA* on π^b is now strongly negative, in stark contrast to its effect on the other two measures. This likely indicates that volatility spikes discourage informed selling, as market makers widen bid-ask spreads and reduce order-book depth.

The results in Columns (3)–(6) indicate that how and to what extent investors use ChatGPT in trading decisions depends on the situation. When faced with positive signals, investors engage in a careful process, evaluating fundamentals, valuation metrics, and qualitative disclosures. This is exactly where ChatGPT excels: summarizing filings, benchmarking peers, and synthesizing scattered information. By using these features, some investors can reduce their informational disadvantage relative to analysts and institutional investors, thereby reducing the buy-side edge of informed traders. In contrast, negative signals seem to trigger a different cognitive response. When facing potential losses, retail investors decide whether to exit immediately or endure volatility. Studies show that such moments activate loss aversion and emotional heuristics (e.g., panic selling, disposition effect, and regret avoidance) (Shefrin and Statman, 1985; Loewenstein, 2000). In these cases, the deliberation process often breaks down for some investors, causing them to ignore analytical tools like ChatGPT and act on instinct. As a result, the advantage that sophisticated traders have on the sell side is likely to stay largely intact.

The way ChatGPT functions can partly be explained by behavioral differences. If its main purpose is to enhance data accessibility or extract proprietary signals, we would expect similar effects on both buying and selling sides. Instead, the stronger reaction to good news indicates that ChatGPT's value lies more in its ability to understand, transforming complex public disclosures into clear insights. It can democratize interpretation (though not necessarily prediction) by helping investors analyze subtle linguistic nuances, uncover hidden information in footnotes, and simplify accounting complexities into more straightforward reasoning. Our findings also align with studies on limited attention (Hirshleifer and Teoh, 2003). Attention is constrained unevenly: investors focus cognitively when spotting opportunities but tune out under stress. Therefore, ChatGPT assists with deliberation when motivation and curiosity are high (with good news), but cannot overcome fear and cognitive overload (with bad news). This pattern resembles Barber and Odean's (2008) discovery that retail investors are more prone to attention-driven buying than disciplined selling. Overall, AI tools help level the playing field when cognitive skills are the main constraint, while having only a modest impact on the psychological barriers that contribute to loss-related concerns and biases.

5.3. ChatGPT's Effects on Price Impact

Next, we analyze the effects of ChatGPT's launch on price-impact parameters, which measure the level of information asymmetry or adverse selection in financial markets. To do this, we estimate two metrics, λ^{GH} and λ^{S}, based on Glosten and Harris (1988) and Sadka (2006). Due to significant kurtosis, the two (raw) measures are winsorized at the 0.5th and 99.5th percentiles (and multiplied by 100 for clarity). In Panel A of Table 4, where λ^{GH} is used as the dependent variable, the coefficient on the interaction term is negative in both columns and significant at the 10% level when we include both fixed effects. In Column (2), the coefficient (-0.0478) on the interaction term indicates that after the release of ChatGPT, the treatment effect on λ^{GH} (a 50.3% drop) is substantial relative to its sample average (0.095). This suggests that compared to the control group of firms, information asymmetry (and thus the adverse-selection cost) in the treatment firms substantially decreases during the post-event period.

The treatment effect is more substantial when we use the Sadka (2006) measure (λ^{S}). Panel B shows that the coefficients are larger in absolute value (-0.1094 and -0.1137) and they are now significant at the 5% level in both columns. This is probably because λ^{S} can better capture the adverse-selection cost by screening out all other irrelevant components (i.e., components related to transitory variable cost, permanent fixed cost, and transitory fixed cost). Note in Eq. (5) in Section 4 that the Sadka model decomposes trading costs into four components (as opposed to two components in the Glosten-Harris model). On balance, the results are consistent with our Hypothesis 3: treatment firms are less subject to information asymmetry and adverse selection after the release of ChatGPT.

Taken together with the results in Table 3, the evidence in Table 4 indicates that ChatGPT's influence propagates through a coherent mechanism: reductions in information asymmetry and adverse selection in the treatment firms eventually manifest as lower informed trading. Lower adverse selection enhances market liquidity and reduces execution costs for uninformed traders, thereby improving overall market quality and potentially lowering firms' cost of capital (Amihud and Mendelson, 1986).

5.4. Analysis of Parallel Trends

Our identification depends on the assumption that, without ChatGPT's release, treatment and control firms would have experienced similar outcome trends. To evaluate the validity of this assumption and to measure the timing and duration of the treatment effect, we employ a dynamic difference-in-differences approach based on Rambachan and Roth (2023). Specifically, we estimate the following regression model: $Y_{i,t} = \sum_{r \neq -1} \beta_r \, TREAT_i + \gamma^T X_{i,t} + \alpha_i + \delta_t + \varepsilon_{i,t}$; where $Y_{i,t}$ denotes the outcome variable and all other variables are the same as in regression model (6).[5] The estimates of β_r capture the dynamic treatment effects for event-time bin r, where $r < 0$ ($r \geq 0$) corresponds to months before (after) the event date of November 30, 2022. Event-time bins are constructed using monthly aggregations. The final pre-event bin ($r = -1$) is omitted and serves as the reference period. When the dependent variable is bounded by [0, 1], we use the fractional logit model. All specifications include firm and date fixed effects, with standard errors clustered two-way at the firm and date levels. Two solid black horizontal lines represent the pre-event and post-event average treatment effects.

Figure 1 shows the event-time coefficients from the dynamic difference-in-differences specification for total informed trading (π^e), informed trading on good news (π^g), informed trading on bad news (π^b), and the two price-impact measures — the Glosten–Harris lambda (λ^{GH}) and the Sadka lambda (λ^S). In the pre-ChatGPT period, although the estimates of β_r in the informed trading regression specification fluctuate over time, there is no evidence of a changing slope or widening gap. Following ChatGPT's release, the β_r estimates in the total informed trading (π^e) model become more negative, driven primarily by informed buying (π^g). Informed selling (π^b) shows little change between the pre- and post-event periods. Both price-impact measures display a similar post-event decline with no pre-event change in slope. The absence of pre-event trend implies that plausible deviations from parallel trends are unlikely to account for the magnitude and persistence of the post-ChatGPT reductions in informed trading and price impact.

5.5. ChatGPT's Effect on Return Volatility

[5] We include the following control variables: *SIZE* (natural log of market capitalization); *BTM* (ratio of book equity to market equity); and *VOLA* (daily standard deviation of intraday returns from WRDS).

Table 5 shows the effects of ChatGPT on return volatility. Panel A reports the results when we measure return volatility by $Vola_{5m}$, i.e., the daily standard deviation of five-minute returns based on transaction prices.[6] Panel B shows the results when we measure return volatility by $Vola_{5m}^{mid}$, i.e., the daily standard deviation of five-minute quote midpoint returns. The results in Table 5 are consistent with Hypothesis 4. The coefficients on (*POST* × *TREAT*) across all four specifications are negative and statistically significant. In Column (1) of Panel B (which uses quote midpoint returns and considers the date fixed effects only), the coefficient (-0.0003) is significant at the 1% level with a *t*-value of -7.44, which is equivalent to an 8.33% decrease in $Vola_{5m}^{mid}$ relative to its daily sample average (0.0036).

When we include joint fixed effects in Column (2), the impact of the interaction term on $Vola_{5m}^{mid}$ is similar. The result with the volatility measure estimated based on observed prices ($Vola_{5m}$) in Panel A is qualitatively similar. These results support the idea that treatment firms with high *CSA* experience better information dissemination and reduced frictions, leading to less price overreaction and lower return volatility post-ChatGPT. This implies that the availability of ChatGPT improves price efficiency in the affected firms in the post-event window. The negative coefficients on firm size indicate that larger firms have lower return volatility, while the mixed signs on book-to-market ratios reflect differences in risk and growth characteristics across firms. Figure 2 shows the results of the parallel trends analysis for the two return volatility measures ($Vola_{5m}$ and $Vola_{5m}^{mid}$). We find that both volatility measures display a similar post-event decline with no pre-event change in slope, satisfying the parallel trends assumption.

Taken together, the findings imply that ChatGPT reduces transient noise in prices—likely by narrowing interpretive differences among investors — thereby stabilizing intraday returns, particularly for firms whose disclosures are more amenable to AI-assisted interpretation. These findings have important implications for policy and regulation. Traditional initiatives (e.g., Reg FD, EDGAR, and XBRL) have aimed to expand access to information. Generative AI introduces a new force that, if guided by proper regulatory frameworks, can go beyond these traditional goals by enabling retail investors to interpret and analyze public disclosures effectively — capabilities that were previously limited to professional analysts or institutional investors.

While some studies suggest that automation and algorithmic coordination can intensify herding or flash-crash phenomena (Kirilenko et al., 2017), generative AI is more likely to have the opposite effect. By enabling diverse investors to independently interpret public information accurately, AI tools like ChatGPT can decrease noise trading and disagreement, thus stabilizing markets. However, we believe that the quality of disclosures remains crucial to realizing these benefits. Our *CSA* framework identifies firms whose detailed, text-heavy filings offer the greatest advantages from AI-assisted interpretation. Consequently,

[6] Given the high correlation between the dependent variable ($Vola_{5m}$ or $Vola_{5m}^{mid}$) and *VOLA*, we exclude *VOLA* as a control variable in the regression.

regulators aiming to harness AI's potential should encourage more comprehensive, structured disclosures that LLMs can process effectively. In an AI-enhanced information environment, simply simplifying language may not suffice for maximizing social welfare.

5.6. Further Analyses: Evidence of Convergence in Trading Activities

In the above section, our results suggest that AI tools may mitigate noise trading and disagreement by helping heterogeneous investors independently arrive at accurate interpretations of public information. In this section, we formally examine whether the availability of ChatGPT induces greater convergence in investors' decision-making processes, drawing on three different order-imbalance measures.

5.6.1. ChatGPT's Release and the Absolute Order Imbalance

We run the (fractional logit) regressions with the daily absolute (dollar-volume-based) order imbalance, defined as $ADOI = \frac{|dbuy-dsell|}{dbuy+dsell}$, which does not distinguish whether it is originated from retail or non-retail investors. We will examine later how the effect changes when we separate the aggregate measure into two different components based on the source of imbalances. The results with *ADOI* are first reported in Columns (1) and (2) of Table 6. Given the relatively higher values of Pseudo R^2 than in Table 3-5, the three firm characteristics (*SIZE*, *BTM*, and *VOLA*) can explain a large portion of the variation in *ADOI*.

Of greater interest is that, regardless of the specifications, the treatment effect remains positive and statistically significant at the 5% level or better in Columns (1) and (2) in Panel A of Table 6. The coefficient on the interaction term varies from 0.0342 to 0.0631 (in log-odds units), with the corresponding *t*-statistics being 4.21 in the first column. As expected, however, the most stringent specification in Column (2) shows that the coefficient (0.0342) on (*POST* × *TREAT*) is smaller than in the other column. Nonetheless, it remains significant at the 5% level, even after controlling for firm characteristics and joint fixed effects. Figure 2 shows the results of the parallel trends analysis for *ADOI*. We find that *ADOI* displays a post-event increase with no pre-event change in slope, satisfying the parallel trends assumption.

Based on the marginal effects of the interaction term reported in Column (1) of Panel B, firms in the treatment group experience an increase of about 0.81 percentage points in absolute order imbalance (*ADOI*) compared to control firms following the release of ChatGPT. Economically, this represents a 5.03% increase relative to the sample mean of *ADOI* (0.161), or approximately 4.50% of its standard deviation (0.180). The effect in Column (2) is slightly smaller but still economically significant. Overall, these results support Hypothesis 5, indicating that the introduction of ChatGPT led to greater convergence in investors' trading decisions, resulting in higher absolute order imbalance for treatment firms. Improved processing of

public information about these firms appears to foster greater consensus among investors in their trading behavior.

The effects of the control variables on *ADOI* are consistent with our expectations. *SIZE* is strongly negatively associated with *ADOI*, indicating that larger firms generally have lower order imbalance, which aligns with the idea that deeper markets attract more diverse investors with different views. Conversely, the impact of *BTM* is positive, possibly because value investors tend to coordinate on contrarian positions when growth stocks dominate market sentiment.

5.6.2. Order Imbalances Originating from Retail vs. Non-Retail Investors

Next, to investigate whether the role of ChatGPT differs across investor types, we decompose the aggregated measure (*ADOI*) into two metrics: the absolute order imbalance stemming from retail investors ($ADOI^{R}$) and that from non-retail (or institutional) investors ($ADOI^{NR}$).

The results with $ADOI^{R}$ in Columns (3) and (4) of Panel A in Table 6 show that the impact of (*POST* × *TREAT*) on $ADOI^{R}$ increases significantly compared to those in Columns (1) and (2): the effects in the corresponding specifications are over 27% larger and significant at the 1% level. Specifically, the coefficient in Column (4) (0.0515) is 50.6% higher. In Panel B, the magnitude of the interaction term's marginal effects is even more noticeable. Across all specifications, its marginal effect exceeds 79% more. Notably, in Column (4), the marginal effect (0.0086) is more than twice that of the corresponding specification in Column (2) of Table 6. The above effect represents a 5.33% increase in $ADOI^{R}$ relative to its sample average (0.272). These results suggest that AI tools can more effectively help retail investors interpret publicly available information about the treatment group of firms during the post-event period.

To examine whether the absolute order imbalance stemming from non-retail investors behaves differently, we now use $ADOI^{NR}$ as the dependent variable in the regressions and present the results in Columns (5) and (6) of Table 6. We find that the interaction term has a positive and significant effect on $ADOI^{NR}$ in both specifications at the 1% level. Notably, however, its marginal effects shown in Panel B are mostly similar to those reported in Columns (1) and (2) of Table 6. That is, while the impact of (*POST* × *TREAT*) on $ADOI^{NR}$ is not trivial, its magnitude is much smaller than when $ADOI^{R}$ is used as the dependent variable. This suggests that ChatGPT helps not only retail investors but also non-retail (or institutional) investors in interpreting the financial statements of treatment firms. However, the effect is considerably stronger for retail investors among the treatment firms, consistent with Hypothesis 6. Figure 2 shows the results of the parallel trends analysis for $ADOI^{R}$ and $ADOI^{NR}$. We find that both measures of order imbalance display a post-event increase with no pre-event change in slope, satisfying the parallel trends assumption.

An additional feature of the results is that ChatGPT's release is also positively associated with order imbalance generated by non-retail investors. This pattern likely reflects two related mechanisms. First,

institutional adoption: a CFA Institute (2023) survey reports that approximately 40% of institutional investors were using ChatGPT by late 2023, suggesting that even sophisticated investors benefit from AI tools when processing complex disclosures. By improving analysts' ability to synthesize information and support portfolio management decisions, ChatGPT can increase alignment in institutional trading. Second, compositional changes in informed trading: as retail investors become better informed through ChatGPT, traditional informed traders — such as hedge funds and analyst-driven strategies — may lose part of their informational advantage. As this advantage erodes, these traders engage in fewer offsetting trades, weakening the countervailing order flow that previously dampened institutional imbalances. As a result, remaining institutional order flows become more coordinated, leading to a higher non-retail order imbalance.

At first glance, the coexistence of increased order imbalances (shown in Table 6) and decreased informed trading (shown in Table 3) seems paradoxical, since classical microstructure models would expect these to move together. However, our findings suggest a nuanced shift in composition that explains this pattern as discussed earlier. Another factor is the democratization effect: ChatGPT allows retail investors to trade more like informed investors, not by gaining private information, but by accurately analyzing public information that was available but too complex to interpret. This creates a new category of "publicly-informed" traders whose coordination mimics informed trading in generating order flows. However, these traders differ because they do not have genuinely private signals. The BHS metric (π^e) is more likely to detect trading based on private information than on public information, though it reflects both. Therefore, publicly-informed retail coordination can lead to higher order imbalances while having little effect on π^e.

In summary, the availability of ChatGPT has influenced trading dynamics in the U.S. stock market by helping diverse, uninformed traders (mostly retail, but some institutional) develop more accurate understandings of public information. This, in turn, leads to greater alignment in trading directions among investors. We believe that this alignment has caused larger order imbalances, with the effect being much more significant for inexperienced retail traders than for institutional traders.

6. Robustness Tests

To test the robustness of our results, we conduct placebo analyses using three pseudo-event dates (May 30, June 30, and November 30, 2021), which occur before ChatGPT's actual release (November 30, 2022). In each case, the pre- and post-event windows are adjusted so that both windows cover a full year around each of these "fake" dates. To conserve space, we provide the method and test results in Internet Appendix B (and Table B1). Here, we briefly summarize the findings.

In Panel A of Table B1, where May 30, 2021 is assumed as the event date, we find that the vast majority of treatment effects are statistically insignificant or have the wrong sign. For example, the three columns that use π^e, π^g, and π^b show that the coefficient on ($POST^{PLB} \times TREAT$) is insignificant. As to the columns that use OLS estimation, the coefficient is positive and significant (with a wrong sign) when $Vola_{5m}$ or $Vola_{5m}^{mid}$ is used. With the three order-imbalance measures, the sign of the coefficient is correct (positive), but none of the three cases is significantly associated with the outcome variable.

Panel B, using June 30, 2021, as the event date, shows that the results are nearly identical. To make the alternative date far from the first one used in Panel A, we select November 30, 2021 as the pseudo-event date in Panel C. Interestingly, the coefficient is significant; however, it again has the opposite sign, indicating that the interaction term is positively related to *ADOI, ADOI*^{R}, and *ADOI*^{NR}. These inconsistent results confirm that our main findings accurately reflect the impact of ChatGPT's release. Specifically, the sign reversals in the order-imbalance and volatility measures indicate that our primary analyses capture real changes in trajectory rather than ongoing trends. Overall, the placebo tests confirm that the treatment and control firms acted similarly before ChatGPT's launch.

Next, we employ an alternative method to identify the treatment group of firms. From the eight Coverage Scores (*CS1*-*CS8*) obtained for each firm, we discard the smallest and largest values to alleviate the potential influence of outliers. We then average the remaining six scores to construct *CSA*^{alt} for each firm. Given this metric (alternative to *CSA*), we classify the sample firms in the top 30% of the *CSA*^{alt} distribution as the treatment group and all others as the control group. The results of DiD analyses using this method are presented in Table B2. In brief, our main findings are minimally affected by this approach. For conciseness, a detailed description is provided in Sub-section B.2. and Table B2 in the Internet Appendix.

7. Does Labor Exposure to Generative AI Moderate the *CSA* Effect?

Our key findings show that generative AI, exemplified by ChatGPT, democratizes information processing, reducing information asymmetry. However, generative AI also influences firm fundamentals (i.e., real economy) by changing production methods and labor structures. Eisfeldt et al. (2025) report significant valuation impacts for firms with workforces highly vulnerable to LLM-driven task substitution, demonstrating another way AI is transforming industries and capital markets. They develop a firm-level metric called "labor/workforce exposure to generative AI" (*LExpo*), which links ChatGPT-assessed occupational task substitutability from the O*NET database with employment data from Revelio Labs. Their research finds that firms with highly automatable tasks experience daily excess returns of about 0.4% in the two weeks following ChatGPT's launch, with effects persisting across industries. These findings

raise several interesting questions. Are the two metrics (*CSA* and *LExpo*) complementary, each capturing a similar aspect of trading behavior, or does *CSA* represent a fundamentally different channel that *LExpo* cannot measure? Do firms at the intersection of both treatment groups (identified by *CSA* and *LExpo*) show stronger effects in reducing information asymmetry in financial markets?

To address the questions above, we examine two ways in which generative AI impacts financial markets. The first is the democratization channel, which works through AI's ability to interpret existing disclosures. High-*CSA* firms produce information-rich statements that ChatGPT can process effectively, reducing the gap between sophisticated and retail investors. The second channel, highlighted by Eisfeldt et al. (2025), relates to firms' operational and labor exposure to AI. High-*LExpo* firms employ workers whose tasks can be automated or augmented, creating strategic uncertainty about future costs, productivity, and business models. In the post-ChatGPT era, with soaring valuations of NVIDIA, Microsoft, Alphabet, Palantir, and other AI-focused companies, financial markets seem to be pricing in the potential for major technological change. Sophisticated investors and analysts, with technical expertise and greater access to company insiders, are better equipped to evaluate these forward-looking narratives. Consequently, high-*LExpo* firms may be more affected by the concepts discussed in "narrative economics" (Shiller, 2017). Meanwhile, retail investors remain at a disadvantage when assessing these narratives.

These two mechanisms can operate simultaneously, either pulling in opposite directions or working together. Firms with high *LExpo* may see a decline in the informativeness of their financial statements, as the traditional links between labor costs, productivity, and profitability weaken. In these firms, narratives and sentiment often dominate market movements, giving institutional investors an edge in assessing future cash flows. Conversely, in firms with low *LExpo* but high *CSA*, retail investors—now better equipped by ChatGPT to analyze accounting statements—continue to trade based on fundamentals. This split may create a market gap: one segment driven by narratives and dominated by informed institutions, and another driven by disclosures where AI helps democratize public information analysis.

7.1. Data on Labor/Workforce Exposure to AI

We obtain the scores for labor exposure to generative AI from Eisfeldt et al. (2025) at the three-digit NAICS level (*LExpo3*) and from Eisfeldt and Schubert (2024) at the two-digit NAICS level (*LExpo2*). These scores, which range from 0 to 1, represent the proportion of tasks or occupations that AI can enhance or replace. We match the sample firms to the scores using Compustat NAICS three-digit codes and augment the metric with two-digit-based scores when three-digit-based scores are unavailable. We use the augmented score (*LExpo*) as the metric for labor exposure to AI.

As shown in Panel A of Table 7, our sample includes 1,927 firms whose score (*LExpo*) is 0.36 on average. The two-digit-based scores (*LExpo2*) are available to more firms (2,236), with a similar mean

value (0.35) and a right-skewed distribution. A smaller standard deviation for *LExpo* (STD: 0.04) indicates that *LExpo* is more centered on its mean than *LExpo2*, given that *LExpo* is constructed from more narrowly defined industries.

Given the labor exposure scores, we identify another treatment group of firms using the 70th percentile of the *LExpo* distribution as a cutoff point, similarly to the case of *CSA*. We then conduct difference-in-differences analyses. Our goal is to estimate regression models that include a three-way interaction term to examine the combined effects of ChatGPT's release on firms' trading behavior at the intersection of two treatment groups in the post-event period.

7.2. LExpo versus CSA

Before moving on to three-way DiD analyses, we first analyze how the two metrics are connected. The correlation coefficients shown in Panel B of Table 7 are interesting: *CSA* and *LExpo* are negatively correlated (-19.9%). Additionally, *CSA* shows a positive correlation (15.2%) with *SIZE*, while *LExpo* is negatively correlated (-3.1%) with it. Moreover, the two metrics are correlated with *BTM* in opposite directions. These results imply that *CSA* and *LExpo* may measure different aspects of effects resulting from ChatGPT's release.

To examine the relation in a multivariate setting, we conduct a panel regression in Panel C where *CSA* serves as the dependent variable and *LExpo* as one of the independent variables. As expected, *LExpo* is negatively related to *CSA* even after controlling for the three firm characteristics. Based on these results, we conclude that firms with highly automatable or replaceable workforces tend to be those for which ChatGPT extracts relatively little insight from their accounting statements. Firms with high *LExpo*, such as IBM (0.488), NVIDIA (0.468), and Microsoft (0.442), employ knowledge workers who perform tasks easily augmented by AI (while replacing workers who perform otherwise). For these firms, strategic uncertainty—whether AI adoption will complement or displace human labor—dominates valuation, and thus traditional accounting variables may offer limited guidance. Conversely, firms with low *LExpo*, such as Walmart (0.235) and McDonald's (0.194), demonstrate stable cost structures and predictable earnings patterns, allowing ChatGPT's textual understanding to provide meaningful analytical value.

These findings motivate a closer examination of firms for which both channels are simultaneously strong. In the next subsection, we therefore focus explicitly on firms at the intersection of high disclosure amenability and high labor exposure to generative AI. This analysis allows us to explore how information democratization and AI-driven real-activity transformation jointly shape trading behavior and price formation when both forces are most salient.

7.3. Regression Results

To examine the joint effects of *CSA* and *LExpo* on our outcome variables, we estimate a regression model similar to Eq. (6) but including the three-way interaction term ($POST \times CSA^H \times LExpo^H$), where CSA^H and $LExpo^H$ equal one for firms in the top 30% of the *CSA* and *LExpo* distributions and zero otherwise. A significantly positive (negative) coefficient on this interaction term would indicate that the post-ChatGPT effect of *CSA* on the outcome variable is stronger (weaker) for firms with high labor exposure to generative AI. Because the model includes both firm and date fixed effects, five terms [*POST*, CSA^H, $LExpo^H$, the intercept, and ($CSA^H \times LExpo^H$)] are fully absorbed and thus not separately identified. Table 8 shows the regression results. The coefficients on the interaction between *POST* and CSA^H remain directionally identical to those reported in earlier tables, indicating that the core *CSA*-based effects continue to hold after controlling for AI labor exposure. However, the coefficients become insignificant for some outcome variables, most notably for total informed trading (π^e) and the Glosten-Harris price-impact measure. This attenuation suggests that while the *CSA* mechanism remains operative, its strength may be reduced in certain dimensions once labor exposure is taken into account.[7]

For most outcome variables, the estimated coefficients on the triple interaction term are not significantly different from zero. These results indicate that, in general, firms' labor exposure to generative AI does not materially alter the post-ChatGPT relation between *CSA* and trading or price-formation outcomes. In other words, the *CSA*-based reduction in information asymmetry and related frictions appears broadly similar across firms with high and low labor exposure to AI. This in turn means that AI's democratization effects are mostly offset by its another role (via the *LExpo* channel.

An important exception is the Sadka (2006) price-impact measure, which shows a positive and statistically significant coefficient on the triple interaction term. This suggests that the beneficial effect on the adverse-selection cost for high-*CSA* firms in the post-ChatGPT period is more than offset when these firms also have high labor exposure to generative AI. A possible explanation is that the Sadka price impact, unlike λ^{GH}, estimates the permanent price effect through unanticipated order flow, making it more closely related to information shocks that update beliefs about future fundamentals rather than predictable trading patterns. Consequently, the Sadka measure (λ^S) may be especially sensitive to forward-looking uncertainty regarding long-term cash flows, technological trends, and business-model changes — areas of valuation that are likely heightened for firms with greater labor exposure to generative AI.

In addition, we show in the previous section (Table 7) that firms with high labor exposure operating in a more narrative-driven economic environment experience a decline in the informativeness of their financial statements, which in turn gives institutional investors a relative advantage in assessing future cash

[7] The reduced significance of the $POST \times CSA^H$ interaction appears to be driven, at least in part, by the three-way DiD specification. Because this specification requires non-missing labor-exposure data, the estimation sample shrinks by roughly 15% in terms of the number of firms. This reduction disproportionately excludes firms with high coverage scores, which attenuates the estimated treatment effect.

flows. The increase in the Sadka price-impact measure may therefore also reflect market makers' response to heightened adverse-selection risk stemming from more informed institutional trading.

8. Conclusion

This study explores how the release of ChatGPT has reshaped the informational structure of the U.S. equity market. By introducing *CSA*, a new metric that measures ChatGPT's ability to interpret firm disclosures, we show that generative AI reduces information asymmetry and changes trading dynamics in meaningful ways. Using a difference-in-differences approach, we find that companies with high *CSA* see significant decreases in informed trading and adverse selection costs, along with lower high-frequency return volatility. These findings suggest that ChatGPT helps make complex disclosures more understandable, narrowing the gap between retail and institutional investors.

Our analyses show that the impact is asymmetric: AI tools have a stronger influence on informed buying than on informed selling, reflecting behavioral selectivity. Additionally, while informed trading decreases, overall order imbalances increase, especially among retail investors, indicating a greater consensus when interpreting public information. These patterns highlight that AI tools enhance market quality not by providing private signals but by enabling better processing of public data. In other words, AI-assisted interpretation can significantly improve retail investors' decision-making without access to private information.

The findings have significant implications for regulation and policy. Traditional efforts like Reg FD and EDGAR focused on access; however, generative AI expands this scope by improving interpretability. Regulators might need to develop disclosure standards that are more detailed and machine-readable to fully harness AI's benefits. Simultaneously, our evidence shows variation: democratization effects are stronger in disclosure-rich, stable industries, while narrative-driven sectors tend to be more affected by information asymmetry. Therefore, AI both equalizes and creates disparities in informational advantages, shaping a market where fundamentals and narratives coexist, often in a divided manner.

To conclude, ChatGPT's release signifies a fundamental shift toward an AI-enhanced information environment. By reducing cognitive barriers and promoting coordinated interpretation of public disclosures, generative AI becomes a powerful driver of transparency and efficiency. These results leave an open question about its long-term balance and interaction with other technological and behavioral factors, as both institutional and retail traders adapt to AI-driven financial markets and economies. Addressing these issues requires further research.

References

Akerlof, G. A. (1970). The market for "lemons": Quality uncertainty and the market mechanism. *Quarterly Journal of Economics* 84, 488–500.

Amihud, Y., and H. Mendelson (1986). Asset pricing and the bid–ask spread. *Journal of Financial Economics* 17, 223–249.

Angrist, J. D., and J.-S. Pischke (2009). *Mostly harmless econometrics: An empiricist's companion*. Princeton, NJ: Princeton University Press.

Barber, B. M., X. Huang, P. Jorion, T. Odean, and C. Schwarz (2024). (Sub)penny for your thoughts: Tracking retail investor activity in TAQ. *Journal of Finance*, forthcoming.

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

Barber, B. M., and T. Odean (2008). All that glitters: The effect of attention and news on the buying of individual and institutional investors. *Review of Financial Studies* 21, 785–818.

Bertrand, M., E. Duflo, and S. Mullainathan (2004). How much should we trust differences-in-differences estimates? *Quarterly Journal of Economics* 119, 249–275.

Boehmer, E., C. M. Jones, X. Zhang, and X. Zhang (2021). Tracking retail investor activity. *Journal of Finance* 76, 2249–2305.

Bloomfield, R. J. (2002). The "Incomplete Revelation Hypothesis" and financial reporting. Working paper.

Brennan, M. J., S.-W. Huh, and A. Subrahmanyam (2018). High-frequency measures of informed trading and corporate announcements. *Review of Financial Studies* 31, 2326–2376.

CFA Institute (2023). *Artificial intelligence in asset management*. Charlottesville, VA: CFA Institute Research Foundation.

Chordia, T., and A. Subrahmanyam (2004). Order imbalance and individual stock returns: Theory and evidence. *Journal of Financial Economics* 72, 485–513.

Chordia, T., R. Roll, and A. Subrahmanyam (2001). Market liquidity and trading activity. *Journal of Finance* 56, 501–530.

Chordia, T., R. Roll, and A. Subrahmanyam (2002). Order imbalance, liquidity, and market returns. *Journal of Financial Economics* 65, 111–130.

Chordia, T., R. Roll, and A. Subrahmanyam (2005). Evidence on the speed of convergence to market efficiency. *Journal of Financial Economics* 76, 271–292.

Chordia, T., R. Roll, and A. Subrahmanyam (2008). Liquidity and market efficiency. *Journal of Financial Economics* 87, 249–268.

Easley, D., N. M. Kiefer, M. O'Hara, and J. B. Paperman (1996). Liquidity, information, and infrequently traded stocks. *Journal of Finance* 51, 1405–1436.

Eisfeldt, A. L., and G. Schubert (2024). Generative AI and finance. Available at SSRN: https://ssrn.com/abstract=4988553 or http://dx.doi.org/10.2139/ssrn.4988553.

Eisfeldt, A. L., G. Schubert, M. B. Zhang, and B. Taska (2025). Generative AI and firm values. *Journal of Finance*, forthcoming.

Eleswarapu, V. R., R. Thompson, and K. Venkataraman (2004). The impact of Regulation Fair Disclosure: Trading costs and information asymmetry. *Journal of Financial and Quantitative Analysis* 39, 209–225.

Feldman, P., J. R. Foulds, and S. Pan (2023). Trapping LLM hallucinations using tagged context prompts. Working paper.

Glosten, L. R., and L. E. Harris (1988). Estimating the components of the bid–ask spread. *Journal of Financial Economics* 21, 123–142.

Glosten, L. R., and P. R. Milgrom (1985). Bid, ask, and transaction prices in a specialist market with heterogeneously informed traders. *Journal of Financial Economics* 14, 71–100.

Hirshleifer, D., and S. H. Teoh (2003). Limited attention, information disclosure, and financial reporting. *Journal of Accounting and Economics* 36, 337–386.

Holden, C. W., and S. Jacobsen (2014). Liquidity measurement problems in fast, competitive markets: Expensive and cheap solutions. *Journal of Finance* 69, 1747–1785.

Hong, H., and J. C. Stein (2007). Disagreement and the stock market. *Journal of Economic Perspectives* 21, 109–128.

Huh, S.-W. (2014). Price impact and asset pricing. *Journal of Financial Markets* 19, 1–38.

Hwang, L., W. Lee, S. Lim, and K. Park (2013). Does information risk affect the implied cost of equity capital? An analysis of PIN and adjusted PIN. *Journal of Accounting and Economics* 55, 148–167.

Imbens, G. W., and T. Lemieux (2008). Regression discontinuity designs: A guide to practice. *Journal of Econometrics* 142, 615–635.

Jones, C. M., and O. A. Lamont (2002). Short-sale constraints and stock returns. *Journal of Financial Economics* 66, 207–239.

Kahneman, D., and A. Tversky (1979). Prospect theory: An analysis of decision under risk. *Econometrica* 47, 263–291.

Kaustia, M. (2010). Prospect theory and the disposition effect. *Journal of Financial and Quantitative Analysis* 45, 791–812.

Kirilenko, A., A. S. Kyle, M. Samadi, and T. Tuzun (2017). The flash crash: High-frequency trading in an electronic market. *Journal of Finance* 72, 967–998.

Kyle, A. S. (1985). Continuous auctions and insider trading. *Econometrica* 53, 1315–1335.

Li, F. (2008). Annual report readability, current earnings, and earnings persistence. *Journal of Accounting and Economics* 45, 221–247.

Li, E. X., and K. Ramesh (2009). Market reaction surrounding the filing of periodic SEC reports. *Accounting Review* 84, 1171–1208.

Loewenstein, G. (2000). Emotions in economic theory and economic behavior. *American Economic Review* 90, 426–432.

Lopez-Lira, A., and Y. Tang (2023). Can ChatGPT forecast stock price movements? Return predictability and large language models. Working paper.

Noy, S., and W. Zhang (2023). Experimental evidence on the productivity effects of generative artificial intelligence. *Science* 381, 187–192.

Rambachan, A., and J. Roth (2023). A more credible approach to parallel trends. *Review of Economic Studies* 90, 2555–2591.

Reuters (2023). *ChatGPT and financial markets*. Thomson Reuters report.

Sadka, R. (2006). Momentum and post-earnings-announcement drift anomalies: The role of liquidity risk. *Journal of Financial Economics* 80, 309–349.

Shefrin, H., and M. Statman (1985). The disposition to sell winners too early and ride losers too long: Theory and evidence. *Journal of Finance* 40, 777–790.

Shiller, R. J. (2017). Narrative economics. *American Economic Review* 107, 967–1004.

Stanco, J., and K. H. Chung (2025). ChatGPT and the stock market. Available at SSRN: https://ssrn.com/abstract=5171287.

Tetlock, P. C. (2007). Giving content to investor sentiment: The role of media in the stock market. *Journal of Finance* 62, 1139–1168.

Wang, S., X. Wei, and X. Zhou (2023). Self-consistency improves chain-of-thought reasoning in language models. *Advances in Neural Information Processing Systems* 36, 35619–35633.

Wei, J., X. Wang, D. Schuurmans, M. Bosma, B. Ichter, F. Xia, E. H. Chi, Q. Le, and D. Zhou (2022). Chain-of-thought prompting elicits reasoning in large language models. Working paper.

Wu, N., M. Gong, L. Shou, S. Liang, and D. Jiang (2023). Large language models are diverse role-players for summarization evaluation. Working paper.

Table 1
Summary Statistics for the Coverage Scores and their Average (*CSA*)

This table reports the summary statistics [mean, median, standard deviation (STD), minimum (Min), 70th percentile (P70), maximum (Max), number of firms (NFirms)] for the Coverage Scores and their average for the sample firms listed on NYSE/AMEX /Nasdaq. The scores are derived by OpenAI's GPT-3.5 Turbo as our LLM model using the Compustat database: i.e., the inputs are each firm's balance sheet and income statement filed in SEC 10K Filing for fiscal year 2022 (the latest pre-ChatGPT year). The Coverage Score (*CS*) is a large language model (LLM)-based metric that quantifies how effectively an uninformed trader could use publicly available financial statements (balance sheet and income statement) with the help of ChatGPT to become more informed and make better trading decisions. To ensure the reliability of *CS*, for each firm we run the same prompt eight times with different temperatures: twice with each of the four different temperatures at 0.0, 0.5, 0.7, and 1.0. Given the eight different *CS*s for each firm, the average score (termed *CSA*) is computed across the eight scores, and this (*CSA*) is used as our key metric to distinguish the treatment group of firms from the control group of firms. The 70th percentile (P70) of the *CSA* distribution (i.e., 58.40) is used as the cut-off point to distinguish the two groups.

Descriptive Statistics for the Coverage Scores

CS by Run	Temperature	Mean	Median	STD	Min	P70	Max	NFirms
CS1	0.0	53.34	60.00	12.79	0.00	60.00	70.00	2867
CS2	0.0	53.30	60.00	12.87	0.00	60.00	70.00	2865
CS3	0.5	51.70	60.00	15.59	0.00	60.00	80.00	2867
CS4	0.5	51.86	60.00	15.48	0.00	60.00	80.00	2867
CS5	0.7	51.25	60.00	16.09	0.00	60.00	85.00	2867
CS6	0.7	51.30	60.00	15.96	0.00	60.00	80.00	2867
CS7	1.0	50.60	50.00	16.78	0.00	60.00	85.00	2866
CS8	1.0	50.79	53.00	16.63	0.00	60.00	90.00	2863
CS Average (CSA)		52.13	55.00	11.39	0.00	58.40	71.67	2867

Table 2
Descriptive Statistics for the Measures Related to Informed Trading and Other Variables

This table reports the descriptive statistics [mean, median, standard deviation (STD), skewness, kurtosis, number of observations in firm-dates (Obs), average number of firms per month (NFirms/mon)] for the measures related to informed trading and price efficiency, as well as other control variables for the sample firms (with share codes 10 and 11) listed on NYSE/AMEX/Nasdaq. To obtain the above items (except for Obs and NFirm/mon), each statistic is first computed for each day, and then the time-series average of daily values is reported. The two-year sample period (from November 30, 2021 to November 29, 2023) includes a pre-event window (11/30/2021-11/29/2022) and a post-event window (12/01/2022-11/29/2023), relative to the event date (November 30, 2022) when ChatGPT was released to the general public. The measures related to informed trading in Panel A are defined as follows. π^e: daily posterior probability of informed trading; π^g (π^b): daily posterior probability of informed buying (selling) on good (bad) news, estimated using daily buyer- and seller-initiated trades following Brennan, Huh, and Subrahmanyam (2018); λ^{GH} and λ^S: price-impact parameter (or the adverse-selection component of trading costs) estimated based on Glosten and Harris (1988) and Sadka (2006), respectively, using intraday dollar order flows within each day. Since the two price-impact measures (λ^{GH} and λ^S) are extremely leptokurtic, they are winsorized at the 0.5th and 99.5th percentiles; and are also multiplied by 10^4, because they have small values. Volatility measures in Pabel B are defined as follows. $Vola_{5m}$: daily standard deviation of intraday returns (using transaction prices) sampled at a five-minute frequency within each trading day; and $Vola_{5m}^{mid}$: daily standard deviation of intraday returns (using bid-ask quote midpoints) sampled at a five-minute frequency within each trading day. Order imbalance measures in Panel C are defined below. *ADOI*: absolute dollar order imbalance computed as $\frac{|dbuy-dsell|}{dbuy+dsell}$, where *dbuy* and *dsell* are daily dollar value of shares bought and that of sold, respectively; $ADOI^R$ ($ADOI^{NR}$): absolute dollar order imbalance caused by retail (non-retail or institutional) investors, for which trades by retail investors are identified following the algorithm proposed by Barber et al. (2024) who improve upon the method developed by Boehmer et al. (2021). To obtain daily buys/sells and quote midpoints, each trade is matched to quotes using the Holden-Jacobsen (2014) algorithm. The three firm characteristics used as control variables in Panel D are obtained as follows. *SIZE*: natural log of market capitalization in $1 [MV: market capitalization (in $100 million) is also shown for interpretational convenience]; *BTM*: ratio of book equity to market equity from Compustat, for which the previous quarter-end book value is matched with daily market-equity values within each quarter; and *VOLA*: daily standard deviation (multiplied by 10^2) of intraday returns from the WRDS Intraday Indicator Database.

	Descriptive Statistics						
	Mean	Median	STD	Skewness	Kurtosis	Obs	NFirms/mon
Variables	**Panel A: Measures Related to Informed Trading**						
π^e	0.3207	0.0687	0.4481	0.8356	-0.9809	784785	1563.32
π^g	0.1573	7.0182E-11	0.3509	2.0253	2.5722	784785	1563.32
π^b	0.1636	6.8315E-13	0.3582	1.9291	2.0997	784193	1562.14
λ^{GH}	0.0949	0.0005	4.5261	1.1070	46.0334	1211119	2412.59
λ^S	0.0465	0.0001	11.2280	0.0874	37.4196	1208132	2406.64
	Panel B: Return Volatility Measures						
$Vola_{5m}$	0.0052	0.0035	0.0047	2.6740	12.5863	1210972	2412.29
$Vola_{5m}^{mid}$	0.0036	0.0027	0.0030	4.2125	40.6243	1210235	2410.83
	Panel C: Absolute Order Imbalances						
ADOI	0.1608	0.1004	0.1798	2.1779	5.2427	1211135	2412.62
$ADOI^R$	0.2719	0.1762	0.2663	1.3368	0.9197	1058666	2108.90
$ADOI^{NR}$	0.1545	0.0930	0.1820	2.3379	5.9666	1068572	2128.63
	Panel D: Control Variables						
SIZE	20.4856	20.5028	2.3499	0.1171	-0.3361	1211135	2412.62
(MV)	119.4629	8.1439	836.1491	22.1555	597.2801	1211135	2412.62
BTM	0.6676	0.5020	0.6989	4.3374	37.5921	1211135	2412.62
VOLA	0.0041	0.0001	0.0280	15.6816	420.3011	1211135	2412.62

Table 3
Effects of ChatGPT's Release on Informed Trading Measures

This table reports the results of difference-in-differences (DiD) analyses to investigate the effects of ChatGPT on informed trading. The two-year sample period includes a one-year window before and after the release of ChatGPT on November 30, 2022, with the pre-event window spanning from 11/30/2021 to 11/29/2022, and the post-event window from 12/01/2022 to 11/29/2023. Panel A contains the results of fractional logit estimation, while Panel B reports the marginal effects (the instantaneous rate of change in the dependent variable with respect to a small change in an explanatory variable) of the variables involved in the interaction term. The dependent variable is one of three measures of informed trading: the daily posterior probability of informed trading (π^e), the daily posterior probability of informed trading on good news (π^g), and the daily posterior probability of informed trading on bad news (π^b) estimated following Brennan, Huh, and Subrahmanyam (2018). The key explanatory variables are defined as follows. *POST*: one if the observation occurs after the event date (11/30/2022), and zero otherwise; *TREAT*: one for firms in the top 30% of the *CSA* distribution, and zero otherwise; and (*POST* x *TREAT*): an interaction term between the above two indicator variables. Control variables include: the natural log of market capitalization (*SIZE*); the book-to-market equity (*BTM*), for which the previous quarter-end book value is matched with daily market-equity values within each quarter to obtain the ratio at a daily frequency; and transaction-based daily return volatility (*VOLA*) obtained from WRDS (multiplied by 10^2). Standard errors are clustered by firm and date, and *t*-statistics are reported in parentheses. The coefficients significantly different from zero at the 10%, 5%, and 1% levels are indicated by *, **, and ***, respectively.

The Effect of ChatGPT on Informed Trading

	Panel A: Fractional Logit Estimation					
	Dep. Var = π^e		**Dep. Var = π^g**		**Dep. Var = π^b**	
Variables	(1)	(2)	(3)	(4)	(5)	(6)
POST x TREAT	-0.0877***	-0.0841**	-0.0808**	-0.0991***	-0.0556*	-0.0326
	(-2.70)	(-2.47)	(-2.38)	(-2.74)	(-1.66)	(-0.94)
TREAT	0.0158		0.0207		0.0011	
	(0.66)		(0.86)		(0.05)	
SIZE	0.0970***	0.1857***	0.0866***	0.3569***	0.0500***	-0.0706**
	(4.60)	(4.55)	(12.55)	(12.14)	(6.33)	(-2.40)
BTM	0.0085	0.0581**	0.0096	0.1480***	0.0020	-0.0436*
	(0.82)	(2.40)	(0.87)	(5.74)	(0.18)	(-1.82)
VOLA	-2.392	-1.081	0.0018	0.0033	-11.33***	-13.43***
	(-0.37)	(-0.16)	(0.10)	(0.31)	(-9.39)	(-9.10)
Firm FE	No	Yes	No	Yes	No	Yes
Date FE	Yes	Yes	Yes	Yes	Yes	Yes
Clustering	Firm & Date	Firm & Date	Firm & Date	Firm & Date	Firm & Date	Firm & Date
Obs	784,785	784,743	784,785	784,781	784,193	784,189
Pseudo R^2	0.0600	0.0866	0.0312	0.0512	0.0267	0.0462
	Panel B: Marginal Effects of the Interaction-term Variables					
POST x TREAT	-0.0180***	-0.0167**	-0.0103**	-0.0124***	-0.0074*	-0.0042
TREAT	0.0032		0.0026		0.0001	

Table 4
Effects of ChatGPT's Release on Price Impact

This table reports the results of difference-in-differences (DiD) analyses to investigate the effects of ChatGPT on price impact. The two-year sample period includes a one-year window before and after the release of ChatGPT on November 30, 2022, with the pre-event window spanning from 11/30/2021 to 11/29/2022, and the post-event window from 12/01/2022 to 11/29/2023. The dependent variable is one of the two the daily price-impact parameters or the adverse-selection components of trading costs: λ^{GH} (used in Panel A), which is estimated based on Glosten and Harris (1988) using intraday dollar order flows available within each day for each stock; and λ^{S} (used in Panel B), which is estimated similarly based on Sadka (2006). Since the estimates of λ^{GH} and λ^{S} are extremely leptokurtic, they are winsorized at the 0.5th and 99.5th percentiles; and are also multiplied by 10^4 (because of small values) before running the regressions. The key explanatory variables are defined as follows. *POST*: one if the observation occurs after the event date (11/30/2022), and zero otherwise; *TREAT*: one for firms in the top 30% of the *CSA* distribution, and zero otherwise; and (*POST* x *TREAT*): an interaction term between the above two indicator variables. Control variables include: the natural log of market capitalization (*SIZE*); the book-to-market equity (*BTM*), for which the previous quarter-end book value is matched with daily market-equity values within each quarter to obtain the ratio at a daily frequency; and transaction-based daily return volatility (*VOLA*) obtained from WRDS (multiplied by 10^2). Standard errors are clustered by firm and date, and *t*-statistics are reported in parentheses. The coefficients significantly different from zero at the 10%, 5%, and 1% levels are indicated by *, **, and ***, respectively.

The Effect of ChatGPT on Price Impact

	Panel A: With Dep. Var. = λ^{GH}		**Panel B: With Dep. Var. = λ^{S}**	
Variables	(1)	(2)	(1)	(2)
POST x TREAT	-0.0441	-0.0478*	-0.1094**	-0.1137**
	(-1.62)	(-1.72)	(-2.00)	(-2.07)
TREAT	0.0115		0.0405	
	(0.65)		(1.36)	
SIZE	-0.0347***	-0.0730***	-0.0125	-0.0775
	(-7.23)	(-3.76)	(-1.15)	(-1.59)
BTM	0.0215	0.0235	0.0354	0.1313
	(1.19)	(1.15)	(0.92)	(1.50)
VOLA	-0.0029	0.0096	0.0154	0.0306
	(-0.30)	(0.92)	(0.66)	(1.29)
Firm FE	No	Yes	No	Yes
Date FE	Yes	Yes	Yes	Yes
Clustering	Firm & Date	Firm & Date	Firm & Date	Firm & Date
Obs	1,211,119	1,211,118	1,208,132	1,208,131
Adjusted R^2	0.0006	0.0083	0.0003	0.0028

Table 5
Effects of ChatGPT's Release on Return Volatility

This table reports the results of difference-in-differences (DiD) analyses to investigate the effects of ChatGPT on return volatility. The two-year sample period includes a one-year window before and after the release of ChatGPT on November 30, 2022, with the pre-event window spanning from 11/30/2021 to 11/29/2022, and the post-event window from 12/01/2022 to 11/29/2023. The dependent variable is one of the two daily return volatility measures: $Vola_{5m}$: daily standard deviation of intraday returns (using transaction prices) sampled at a five-minute frequency within each trading day; and $Vola_{5m}^{mid}$: daily standard deviation of intraday returns (using bid-ask quote midpoints) sampled at a five-minute frequency within each trading day. To obtain daily buys/sells and quote midpoints, each trade is matched to quotes using the Holden-Jacobsen (2014) algorithm. Panel A contains the result with $Vola_{5m}$, while panel B does the same with $Vola_{5m}^{mid}$. The key explanatory variables are defined as follows. *POST*: one if the observation occurs after the event date (11/30/2022), and zero otherwise; *TREAT*: one for firms in the top 30% of the *CSA* distribution, and zero otherwise; and (*POST* x *TREAT*): an interaction term between the above two indicator variables. Control variables include: the natural log of market capitalization (*SIZE*); the book-to-market equity (*BTM*), for which the previous quarter-end book value is matched with daily market-equity values within each quarter to obtain the ratio at a daily frequency. Given the high correlation between the dependent variable and *VOLA*, the latter is not included in the regressions. The coefficients significantly different from zero at the 10%, 5%, and 1% levels are indicated by *, **, and ***, respectively.

The Effect of ChatGPT on Return Volatility

	Panel A: With Dep. Var. = $Vola_{5m}$		**Panel B: With Dep. Var. = $Vola_{5m}^{mid}$**	
Variables	(1)	(2)	(1)	(2)
POST x TREAT	-0.0004***	-0.0002***	-0.0003***	-0.0002***
	(-5.66)	(-3.27)	(-7.44)	(-5.46)
TREAT	0.0001		0.0001*	
	(1.13)		(1.74)	
SIZE	-0.0013***	-0.0020***	-0.0006***	-0.0009***
	(-46.76)	(-27.47)	(-42.86)	(-19.03)
BTM	-0.0003***	0.0002***	-0.0003***	-0.00001**
	(-4.12)	(3.27)	(-6.65)	(-1.96)
Firm FE	No	Yes	No	Yes
Date FE	Yes	Yes	Yes	Yes
Clustering	Firm & Date	Firm & Date	Firm & Date	Firm & Date
Obs	1,210,972	1,210,971	1,210,235	1,210,234
Adjusted R^2	0.4234	0.6368	0.2317	0.4300

Table 6
Effects of ChatGPT's Release on Absolute Order Imbalances

This table reports the results of difference-in-differences (DiD) analyses to investigate the effects of ChatGPT on absolute order imbalance. The two-year sample period includes a one-year window before and after the release of ChatGPT on November 30, 2022, with the pre-event window spanning from 11/30/2021 to 11/29/2022, and the post-event window from 12/01/2022 to 11/29/2023. Panel A contains the results of fractional logit estimation, while Panel B reports the marginal effects (the instantaneous rate of change in the dependent variable with respect to a small change in an explanatory variable) of the variables involved in the interaction term. The dependent variable is one of three measures of order imbalance: the daily absolute (dollar-volume-based) order imbalance defined as $ADOI = \frac{|dbuy-dsell|}{dbuy+dsell}$, where *dbuy* and *dsell* are daily dollar value of shares bought and that of sold, respectively, the daily absolute (dollar-volume-based) order imbalance caused by retail investors, denoted as $ADOI^R$ and daily absolute (dollar-volume-based) order imbalance caused by retail investors, denoted as $ADOI^{NR}$. We identify retail vs. non-retail trades in the TAQ database by using the algorithm proposed by Barber et al. (2024) who improve upon the method developed by Boehmer et al. (2021). The key explanatory variables are defined as follows. *POST*: one if the observation occurs after the event date (11/30/2022), and zero otherwise; *TREAT*: one for firms in the top 30% of the *CSA* distribution, and zero otherwise; and (*POST* x *TREAT*): an interaction term between the above two indicator variables. Control variables include: the natural log of market capitalization (*SIZE*); the book-to-market equity (*BTM*), for which the previous quarter-end book value is matched with daily market-equity values within each quarter to obtain the ratio at a daily frequency; and transaction-based daily return volatility (*VOLA*) obtained from WRDS (multiplied by 10^2). Standard errors are clustered by firm and date, and *t*-statistics are reported in parentheses. The coefficients significantly different from zero at the 10%, 5%, and 1% levels are indicated by *, **, and ***, respectively.

The Effect of ChatGPT on Absolute Order Imbalance

	Panel A: Fractional Logit Estimation					
	Dep. Var = *ADOI*		**Dep. Var = $ADOI^R$**		**Dep. Var = $ADOI^{NR}$**	
Variables	(1)	(2)	(1)	(2)	(1)	(2)
POST x TREAT	0.0631***	0.0342**	0.0807***	0.0515***	0.0729***	0.0398***
	(4.21)	(2.50)	(5.17)	(3.53)	(4.39)	(2.65)
TREAT	-0.1172***		-0.1284***		-0.1088***	
	(-4.78)		(-5.16)		(-4.42)	
SIZE	-0.2578***	-0.1832***	-0.2843***	-0.1711***	-0.3098***	-0.2518***
	(-22.42)	(-15.02)	(-47.23)	(-11.33)	(-25.18)	(-16.67)
BTM	0.0300**	0.0493***	0.0979***	0.0680***	0.0485***	0.0542***
	(2.38)	(5.48)	(5.07)	(5.96)	(3.73)	(4.27)
VOLA	2.428	0.0008	14.90***	5.626***	3.129	0.0003
	(0.73)	(0.08)	(11.22)	(6.72)	(0.79)	(0.04)
Firm FE	No	Yes	No	Yes	No	Yes
Date FE	Yes	Yes	Yes	Yes	Yes	Yes
Clustering	Firm & Date	Firm & Date	Firm & Date	Firm & Date	Firm & Date	Firm & Date
Obs	1,211,135	1,211,134	1,125,607	1,125,605	1,136,002	1,136,000
Pseudo R^2	0.2252	0.4093	0.2669	0.4545	0.2707	0.4317
	Panel B: Marginal Effects of the Interaction-term Variables					
POST x TREAT	0.0081***	0.0042**	0.0145***	0.0086***	0.0090***	0.0046***
TREAT	-0.0150		-0.0230***		-0.0134***	

Table 7

Relationship between *CSA* and the Labor Exposure Score (*LExpo*)

This table reports summary statistics [mean, standard deviation (STD), median, minimum (Min), maximum (Max), number of firms (NFirms)] for the labor exposure score and its relationship to *CSA*. The score of labor exposure to AI (*LExpo*) measures the extent to which labor tasks in each industry (based on NAICS) are subject to automation or augmentation by generative AI. The score is derived from O*NET task data and LLM-based exposure ratings and aggregated at the three- or two-digit level of NAICS codes. The three- and two-digit-based labor exposure scores are denoted as *LExpo3* and *LExpo2*, respectively. *LExpo3* is obtained from Eisfeldt, Schubert, Zhang, and Taska (2023), which provides greater industry granularity; and *LExpo2* is obtained from Eisfeldt and Schubert (2024). *LExpo* used in Panels A-C is *LExpo3* augmented by using *LExpo2* for industries where *LExpo3* is not available, in order to preserve more sample firms. Panel A contains the summary statistics for the two labor exposure scores. Panels B and C report the correlation coefficients between variables, and the results of panel regressions to examine the relationship between firms' labor exposure to generative AI (*LExpo3*) and *CSA*. *CSA* is our LLM-based metric that quantifies how effectively an uninformed trader can use publicly available financial statements (balance sheet and income statement) with the help of ChatGPT to become more informed and make better trading decisions. The sample period is from November 30, 2021 to November 29, 2023. The regression specification for Panel C is: $CSA_{i,t} = a + \varphi_1 LExpo_i + \varphi_2 SIZE_{i,t} + \varphi_3 BTM_{i,t} + \varphi_4 VOLA_{i,t} + \varepsilon_i$, where *t*-statistics are reported in parentheses. The coefficients significantly different from zero at the 10%, 5%, and 1% levels are indicated by *, **, and ***, respectively.

Panel A: Descriptive Statistics for the Labor Exposure Scores

Variables	Mean	STD	Median	Min	Max	NFirms
LExpo	0.36	0.04	0.36	0.24	0.42	1927
LExpo2	0.35	0.13	0.24	0.08	0.49	2236

Panel B: Correlations between Variables

	CSA	LExpo	SIZE	BTM	VOLA
CSA	1.000				
LExpo	-0.199	1.000			
SIZE	0.152	-0.031	1.000		
BTM	-0.133	0.011	-0.360	1.000	
VOLA	-0.006	0.002	-0.025	0.009	1.000

Variables	**Panel C: Panel Regression Result: Dep. Var. = CSA**
LExpo	-29.9693***
	(-204.81)
SIZE	0.3240***
	(77.86)
BTM	-1.3928***
	(-103.37)
VOLA	-0.0794**
	(-2.11)
Constant	58.3921***
	(551.30)
Obs	1,034,308
Adjusted R^2	0.0631

Table 8

Effects of ChatGPT's Release: Information Democratization vs. Value Enhancing via Labor Exposure to AI

This table reports the results of difference-in-differences (DiD) analyses to investigate whether the intersection of high *CSA* and high *LExpo* can better capture informed trading activities after ChatGPT's release. The two-year sample period includes a one-year window before and after the release of ChatGPT on November 30, 2022, with the pre-event window spanning from 11/30/2021 to 11/29/2022, and the post-event window from 12/01/2022 to 11/29/2023. The two metrics to identify treatment firms (*CSA* and *LExpo*) are defined in Tables 1 and 11, respectively. Most of the key variables are defined as in the previous tables. The remaining are defined as follows. *POST*: one if the observation occurs after the event date, and zero otherwise; CSA^H: one for firms in the top 30% of the *CSA* distribution, and zero otherwise; $LExpo^H$: one for firms in the top 30% of the *LExpo* distribution, and zero otherwise; and (*POST* x CSA^H x $LExpo^H$): an interaction term between the three variables. Note that the four variables (*POST*, CSA^H, $LExpo^H$, and the intercept) as well as one of the three 2-way interaction terms (CSA^H x $LExpo^H$) are absorbed by including both firm and date fixed effects. When the dependent variable is bounded by [0, 1], fractional logit estimation is used, instead of OLS estimation. In that case, marginal effects are not reported for brevity. Standard errors are clustered by firm and date, and *t*-statistics are reported in parentheses. The coefficients significantly different from zero at the 10%, 5%, and 1% levels are indicated by *, **, and ***, respectively.

Heterogeneous Effects of ChatGPT: Information Democratization vs. Value Enhancing through Labor Exposure

Estimation Method:	Fractional Logit			OLS				Fractional Logit		
Dep. Var.:	π^e	π^g	π^b	λ^{GH}	λ^S	$Vola_{5m}$	$Vola_{5m}^{mid}$	*ADOI*	$ADOI^R$	$ADOI^{NR}$
Variables	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)
POST x CSA^H x $LExpo^H$	-0.0045	0.0605	-0.0726	0.0106	0.2105*	-0.0001	0.0000	-0.0473	0.0086	-0.0188
	(-0.06)	(0.71)	(-0.96)	(0.16)	(1.86)	(-0.50)	(0.67)	(-1.48)	(0.23)	(-0.53)
POST x CSA^H	-0.0532	-0.0916**	0.0101	-0.0542	-0.0169**	-0.0002**	-0.0003***	0.0605***	0.0615***	0.0577***
	(-1.27)	(-2.04)	(0.24)	(-1.39)	(-2.28)	(-2.66)	(-6.59)	(3.46)	(3.60)	(3.10)
POST x $LExpo^H$	0.0211	0.0034	0.0379	-0.0304	-0.1243	-0.0002**	-0.0003***	0.0760***	0.0459**	0.0717***
	(0.47)	(0.07)	(0.85)	(-0.79)	(-1.55)	(-2.07)	(-5.08)	(3.76)	(2.31)	(3.51)
SIZE	0.1835***	0.3621***	-0.0848***	-0.0748***	-0.0918*	-0.0020***	-0.0009***	-0.1853***	-0.1705***	-0.2506***
	(4.84)	(11.01)	(-2.69)	(-3.53)	(-1.66)	(-25.00)	(-17.24)	(-13.83)	(-10.59)	(-16.49)
BTM	0.0574**	0.1469***	-0.0442*	0.0275	0.1137	0.0001	-0.0001**	0.0487***	0.0735***	0.0558***
	(2.30)	(5.32)	(-1.77)	(1.17)	(1.35)	(1.39)	(-2.12)	(4.87)	(5.98)	(4.13)
VOLA	-0.1937	0.0048	-13.1000***	0.0114	0.0299			-0.0006	5.2670***	-0.0007
	(-0.05)	(0.55)	(-8.33)	(0.98)	(1.31)			(-0.07)	(5.64)	(-0.11)
Firm FE	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Date FE	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Clustering	Firm & Date	Firm & Date	Firm & Date	Firm & Date	Firm & Date	Firm & Date	Firm & Date	Firm & Date	Firm & Date	Firm & Date
Obs	679,730	679,768	679,252	1,034,291	1,031,687	1,034,291	1,031,687	1,034,307	962,043	970,631
Pseudo/Adjusted R^2	0.0866	0.0518	0.0461	0.0088	0.0028	0.6315	0.4272	0.4066	0.4504	0.4283

Figure 1. Analysis of parallel trends in informed trading and trade price impact

To evaluate the validity of the parallel trends assumption, we employ a dynamic difference-in-differences approach based on Rambachan and Roth (2023). Specifically, we estimate: $Y_{i,t} = \sum_{r \neq -1} \beta_r \, TREAT_i + \gamma^T X_{i,t} + \alpha_i + \delta_t + \varepsilon_{i,t}$; where $Y_{i,t}$ denotes the following outcome variables: total informed trading (π^e), informed trading on good news (π^g), informed trading on bad news (π^b), Glosten–Harris lambda (λ^{GH}), and Sadka lambda (λ^S). The estimates of β_r capture the dynamic treatment effects for event-time bin r, where $r < 0$ ($r \geq 0$) corresponds to months before (after) the event date of November 30, 2022. Event-time bins are constructed using monthly aggregations. The final pre-event bin ($r = -1$) is omitted and serves as the reference period. When the dependent variable is bounded by [0, 1], we use the fractional logit model. All specifications include firm and date fixed effects, with standard errors clustered two-way at the firm and date levels. Two solid black horizontal lines represent the pre-event and post-event average treatment effects.

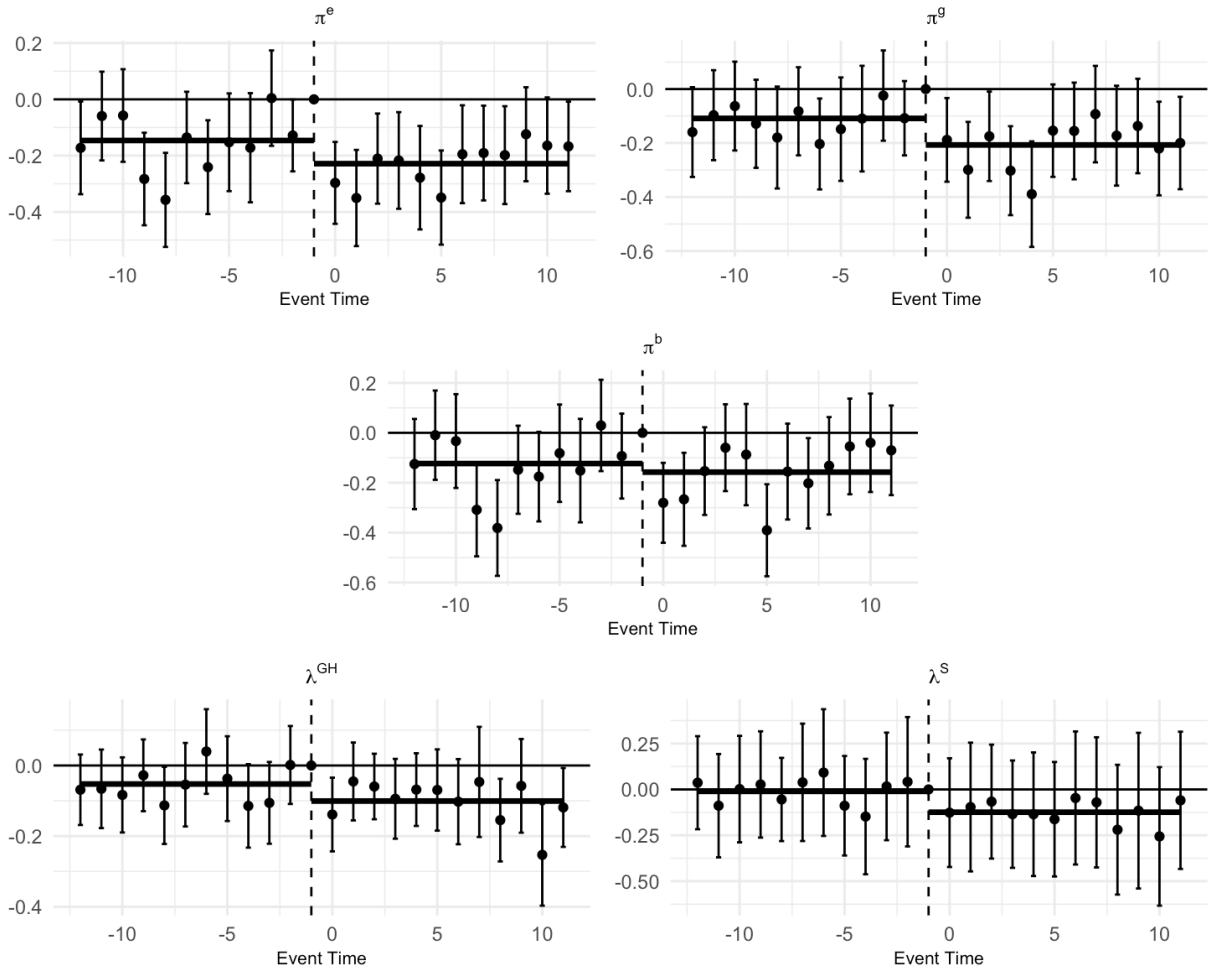
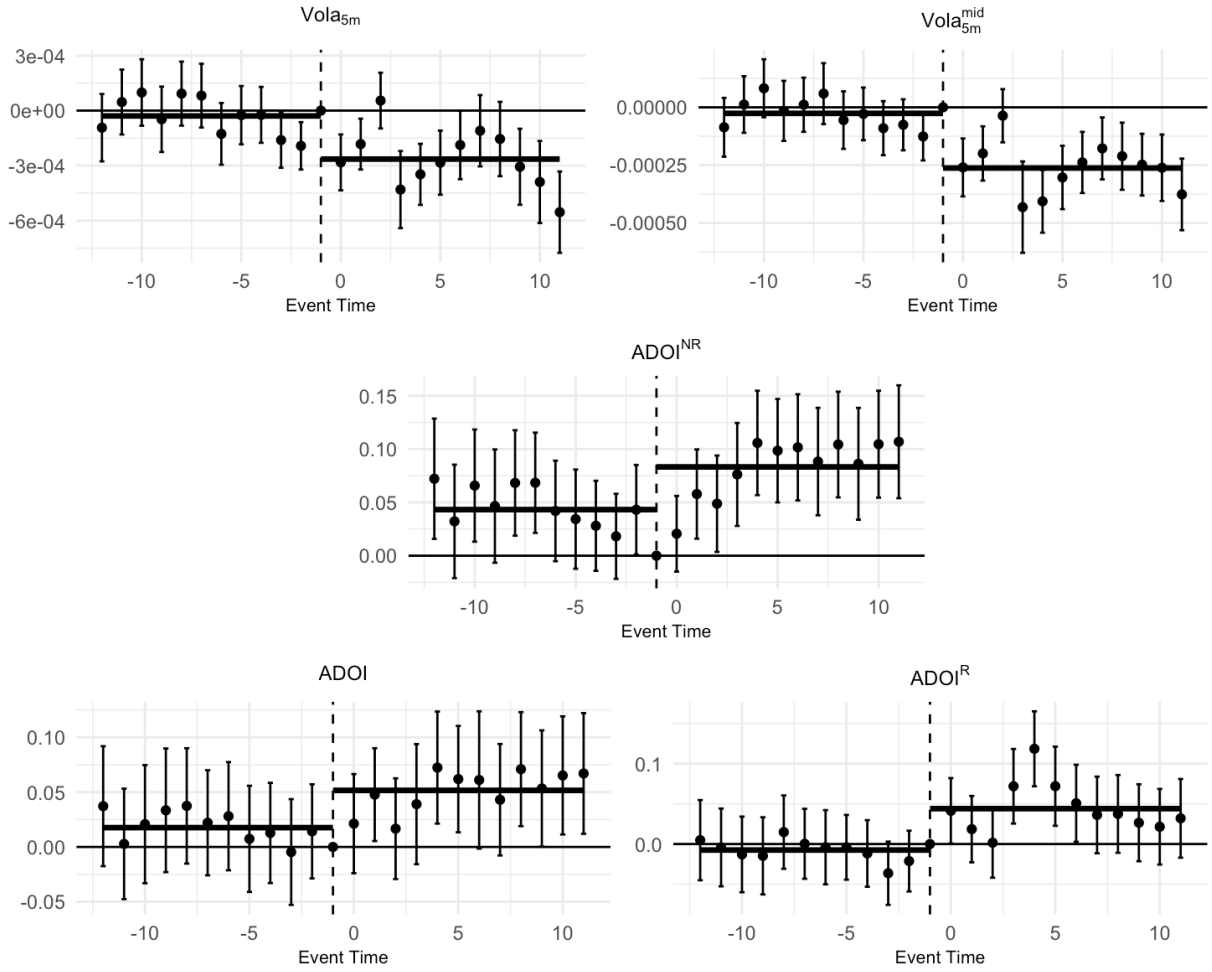


Figure 2. Analysis of parallel trends in return volatility and order imbalance

To evaluate the validity of the parallel trends assumption, we employ a dynamic difference-in-differences approach based on Rambachan and Roth (2023). Specifically, we estimate: $Y_{i,t} = \sum_{r \neq -1} \beta_r \, TREAT_i + \gamma^T X_{i,t} + \alpha_i + \delta_t + \varepsilon_{i,t}$; where $Y_{i,t}$ denotes the following outcome variables: the daily standard deviation of five-minute returns based on transaction prices ($Vola_{5m}$), the daily standard deviation of five-minute quote midpoint returns ($Vola_{5m}^{mid}$), the total absolute order imbalance (*ADOI*), the absolute order imbalance stemming from retail investors ($ADOI^R$), and the absolute order imbalance stemming from non-retail investors ($ADOI^{NR}$). The estimates of β_r capture the dynamic treatment effects for event-time bin r, where $r < 0$ ($r \geq 0$) corresponds to months before (after) the event date of November 30, 2022. Event-time bins are constructed using monthly aggregations. The final pre-event bin ($r = -1$) is omitted and serves as the reference period. When the dependent variable is bounded by [0, 1], we use the fractional logit model. All specifications include firm and date fixed effects, with standard errors clustered two-way at the firm and date levels. Two solid black horizontal lines represent the pre-event and post-event average treatment effects.



INTERNET APPENDIX A: Methodology for Deriving the Coverage Score (*CS*) and its Average (*CSA*)

A.1. Data Inputs and Standardization

To get the Coverage Score (*CS*) for each firm, we draw on Compustat for firms' fundamental data. Specifically, we extract each firm's annual SEC 10-K filing for 2022 (the latest pre-ChatGPT year), which contains balance sheets for 2021–2022 and income statements for 2020–2022. We choose this year for several reasons: (a) 2022 annual reports align with the timeframe when ChatGPT was first released to the public; (b) using a single annual filing's data ensures a consistent basis for comparison across firms and avoids needing the model to parse multi-year trends; and (c) by limiting the timeframe to the year immediately before ChatGPT was widely adopted, we minimize confounding effects that might be caused by the release of newer versions of ChatGPT or changing data regimes. In essence, data in 2022's annual filings serve as a "snapshot" of firm fundamentals at the dawn of the LLM era.

We query the Compustat Fundamentals database to retrieve all relevant fields for each firm in 2022. We include all firms with share codes 10 and 11 in CRSP. After retrieving the raw data, we standardize to make them suitable for the LLM. First, we address the issue of financial statement line items, which often have different, firm-specific names or titles. We map each raw Compustat field to a common descriptive name (e.g., "che" to "cash and short-term investments") and ensure consistency across firms. As the benchmark list of items, we make a comprehensive dictionary of balance sheet and income statement items (over 100 items each: see how to design the prompt below). Each firm's data are aligned to this list.

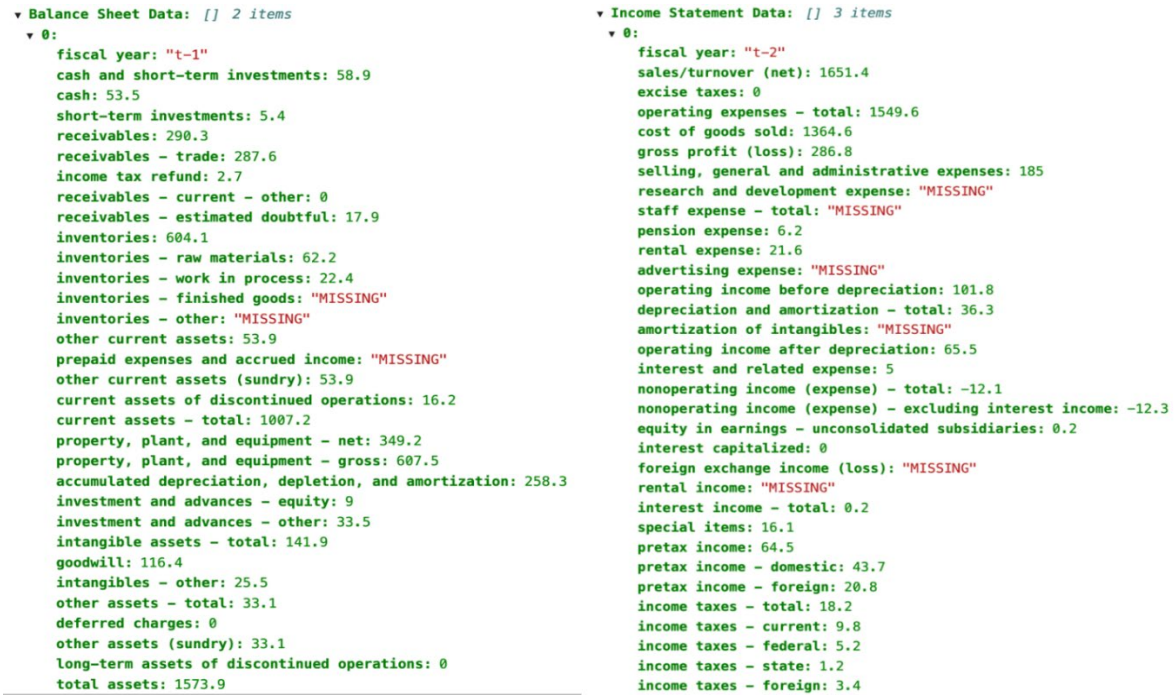
Next, if a firm does not report a particular item, we explicitly mark it as missing. Essentially, for each firm we construct a comprehensive schema encompassing all potential items, where values are entered when available and left blank otherwise. This lets the LLM recognize not only what item exists, but also what is absent. Highlighting missing pieces is crucial, since a key part of *CS* is whether the statements are complete or sparse. Another important feature of *CS* is that it distinguishes between zeros and missing values, thereby capturing firms' voluntary disclosure choices. Some companies disclose line items as zero when they are irrelevant to operations, providing clarity to investors. But others omit the items themselves altogether, in which case we mark the items as missing. From the perspective of uninformed investors, this distinction is meaningful: explicit zeros reduce ambiguity and make statements more usable, whereas omissions create uncertainty. Accordingly, a lower *CS* value due to the presence of missing items does not constitute a penalty for lacking certain activities but rather reflects insufficient disclosure. This design ensures that the *CS* metric consistently measures the amount of usable information that firms actually provide, independent of whether they engage in specific activities.

To prevent the model or platform (OpenAI's GPT-3.5 Turbo) from leveraging external real-time information associated with specific fiscal years (e.g., "Apple 2022" news or market data), we anonymize the periods. For the 2022 financial statement set, we include the 2022 and 2021 balance sheets, along with

the 2022, 2021, and 2020 income statements.[1] Instead of supplying the explicit years, we relabel them relative to the most recent period: i.e., the current year with *t*, the prior year *t–1*, and the second prior year *t–2*. This ensures that the model cannot cross-reference the years with external knowledge sources and remains focused solely on the internal structure and relationships within the financial statements when generating *CS*. This structured format is fed directly into the prompt (input text or instruction), following the list of all possible items (to indicate what could have been reported). That way, the model can parse the information consistently, utilizing the standardized format working as a fact sheet. Figure A1 below illustrates some parts of the finally processed financial statements, which serve as an input to the LLM.

A related concern in LLM-based analyses is whether model outputs might reflect external or post-event information. This concern does not apply here for two reasons. First, OpenAI's GPT-3.5 Turbo has a documented training cutoff in May 2022, while the 2022 Form 10-K filings used in our analysis became publicly available mainly in early 2023—well after the model's training ended. Second, our prompt design adds extra safeguards: inputs include only structured financial statement data and exclude firm identifiers, realized outcomes, market reactions, and forward-looking information. Collectively, these features ensure that the Coverage Score reflects disclosure structure and completeness rather than post-event knowledge.

Figure A1: Sample Items of a Firm's Standardized Balance Sheet and Income Statement



[1] For the annual 10-K filings, firms are required by the SEC to present comparative financial information. That is, a firm's 10-K report (and hence Compustat data) for any fiscal year includes balance sheets for two years and income statements for three years.

A.2. Prompt Design

Designing the prompt is an iterative process, informed by best practices in prompt engineering and careful consideration of our research question. As the final step, our prompt design aims to simulate ChatGPT assisting an uninformed retail trader who examines a particular company's financial statements. The prompt explicitly defines the role and context, provides the relevant data, and instructs the model on how to evaluate that data and produce the coverage score.

We begin with a system message (in the OpenAI API format) or a preamble: "You are ChatGPT assisting an uninformed trader who has no accounting background and limited finance knowledge." This establishes the AI's persona as a helpful explainer to a novice. Next, by dictating, "… whether this company's publicly disclosed financial statements are sufficient to make a basic trading decision (e.g., BUY/SELL/AVOID)," we clarify that the trader's goal is to decide whether the firm's disclosed statements are sufficient enough to make a trading decision. This context is crucial: in our study, the model is *not* asked to predict returns or to give complex recommendations, but rather to *assess* whether the available data is adequate for making an informed decision. Simply put, the question we ask is: Is there enough information for a naive investor (with AI's help) to form a reasonable bullish or bearish inclination? The key parts of the prompt then specify that the model is given two sets of information as follows:

- "1. The full list of possible financial variables available from Compustat." This refers to the exhaustive list of balance sheet and income statement items. By presenting the full menu of items (over 100 for each statement category), we set a benchmark for completeness. The model can see, for instance, that a company could report items such as R&D expense, various tax items, and interest expense, etc., and thus if many of those are missing from the actual data, it will recognize that the coverage is partial. We include these lists in the 'JSON' array format within the prompt. This essentially serves as a checklist and makes the inputs explicit and structured for proper reasoning by the model.
- "2. The company's actual reported data, which may be incomplete." Following this message, we embed the firm-specific data under a clear label. This is the primary input that the model will analyze. By structuring it under "Balance Sheet Data" and "Income Statement Data," we make it easy for ChatGPT to discuss each financial statement separately if needed. Providing a well-labeled, schema-like structure aligns with Wei et al.'s (2022) suggestion to reduce ambiguity.

With the above, the model has all the raw material it needs. The remainder of the prompt focuses on what the model should do with the materials provided. We explicitly instruct the AI to simulate how well it can help the trader interpret the financials, by focusing on whether there is enough data to assess the company's profitability, solvency, and basic risk profile. Each step in the prompt is enumerated (detailed in Figure A2 below), and its key points are described below:

- "1. Analyze the data and judge whether it is informative for key financial questions." By this statement, our purpose is to ask: e.g., Can we infer if the firm is profitable? Is it growing? Are its

debt levels a cause for concern? If crucial items (such as revenue or liabilities) are missing, it signals a red flag.

- "2. Note whether either the balance sheet or income statement is missing or very sparse." If the company didn't report a balance sheet at all (which can happen for some firms), or if many line items are null, the model should reflect this information on its *CS* accordingly.
- "3. Estimate how much a naive trader could realistically learn from the data with ChatGPT's help." This instruction encourages the model to put itself in the shoes of an uninformed trader and assess the usability of a firm's data.
- "4. Assign a numeric Coverage Score from 0 to 100, …, 50 points = coverage of data (how complete it is), 50 points = usability for decision-making (how helpful ChatGPT could be with this data)." In this step, we require the model to report the final score in a specific format: it must generate an output as in "Coverage Score: X," where X is the value of *CS*. This standardized output makes it easy for us to parse the model's response files.

Breaking the task into explicit, auditable sub-steps is consistent with the prompt-design guidance for structuring queries to improve reliability and interpretability. We emphasize in the prompt that the model should produce the analysis first, and then the score at the very end. This way we capture the model's chain of thoughts and justifications, not just the score. In fact, the intermediate analysis is useful for validation, and it often reveals why the model assigns a particular score (e.g., "the income statement is missing key expense details, limiting insight into profitability, so I can only give a moderate score.").

Figure A2: Detailed Instructions in the Prompt

```
prompt <- paste(
  "You are ChatGPT assisting an uninformed trader who has no accounting background and limited finance knowledge.",
  "The trader wants to understand whether this company's publicly disclosed financial statements are sufficient to make a basic trading decision (e.g., BUY / SELL / AVOID).",
  "",
  "You are provided with:",
  "1. The full list of possible financial variables available from Compustat.",
  "2. The company's actual reported data, which may be incomplete.",
  "",
  "Your task is to simulate how well **you (ChatGPT)** can help the trader interpret the financials.",
  "Your analysis should focus on whether there is enough data to:",
  "• Understand the company's profitability, solvency, and basic risk.",
  "• Make a **simple, reasonable trading decision** (e.g., 'buy if growth seems healthy', 'avoid if key data is missing', etc.).",
  "",
  "**Evaluation Instructions:**",
  "1. Analyze the data and judge whether it is informative for key financial questions.",
  "2. Note whether the balance sheet or income statement is missing or sparse.",
  "3. Estimate how much a naive trader could realistically learn from this data with your help.",
  "4. Assign a numeric **Coverage Score** from 0 to 100, where:",
  "    - 50 points = coverage of data (how complete it is)",
  "    - 50 points = usability for decision-making (how helpful ChatGPT could be with this data)",
  "",
  "**Final Output (Required):**",
  "• Please follow this exact format: Coverage Score: [your score]",
  "",
  "Here is the benchmark list of Balance Sheet variables:",
  toJSON(max_balance_vars, auto_unbox = TRUE),
  "",
  "Here is the benchmark list of Income Statement variables:",
  toJSON(max_income_vars, auto_unbox = TRUE),
  "",
  "Here is the company's reported data:",
  toJSON(list(
    'Balance Sheet Data' = balance_sheet,
    'Income Statement Data' = income_statement), auto_unbox = TRUE)
)

return(prompt)
}
```

A.3. Rationale for the Design and the Choice of a Model

We instruct the model to behave from an uninformed trader's perspective rather than a forensic accountant. This reflects our efforts to mimic a realistic case: a retail investor seeking help from ChatGPT. To align with that persona, the prompt uses simple, accessible language rather than technical jargon. By positioning the AI as a novice-friendly guide, the model's outputs can better assist the intended audience.

We also deliberately avoid constraining the model's response length or format beyond the final score line. Instead of restricting it to one-sentence answers or purely numerical outputs, we required a free-form analysis followed by the coverage score. This structure allows the model to engage in self-reflection, effectively 'thinking out loud.' Studies in the LLM literature follow a similar approach: allowing models to generate intermediate reasoning (known as Chain-of-Thought prompting) leads to more accurate and interpretable results (Wei et al., 2022). In our case, having the model articulate its reasoning ensures that it considers both the presence or absence of financial data and its usefulness before assigning a score. This also improves transparency, as we can check the likelihood of errors or inconsistencies.

The finalized prompt shown in Figure A2 above is carefully structured to avoid biasing the model toward any particular score. We do not include sample scores that could anchor its outputs. Each firm is evaluated independently, with the only guidance provided by the scoring framework: a range from 0 to 100 with equal weights given to coverage and usability. To obtain the finalized version, we refined the prompt through multiple trial runs to ensure it produces focused and sensible responses. Early versions were adjusted when the model strayed off-topic or neglected to address the issue of missing data.

Figure A3: Prompt-guided ChatGPT Output for Sterling Infrastructure, Inc. (STRL)

```
Coverage Score: 70

The provided financial data offers a comprehensive view of the company's financial position and performance. Here
are some key points to consider:

### Balance Sheet Analysis:
- **Completeness:** The balance sheet data is quite detailed, providing insights into the company's assets,
liabilities, and equity. However, there are missing values in several key areas like short-term borrowings, debt
maturing in the 2nd and 3rd year, and certain debt types.
- **Usability:** With the available data, you can evaluate the company's liquidity, solvency, and capital structure.
Key metrics like total assets, total liabilities, and equity are present, enabling basic analysis of the company's
financial health.

### Income Statement Analysis:
- **Completeness:** The income statement data is mostly complete, offering a look into the company's revenue,
expenses, and net income. However, there are missing values for some expenses like research and development, staff
expenses, and certain gains/losses.
- **Usability:** You can assess the company's revenue generation, profitability, and operational efficiency based on
metrics like gross profit, operating income, and net income.

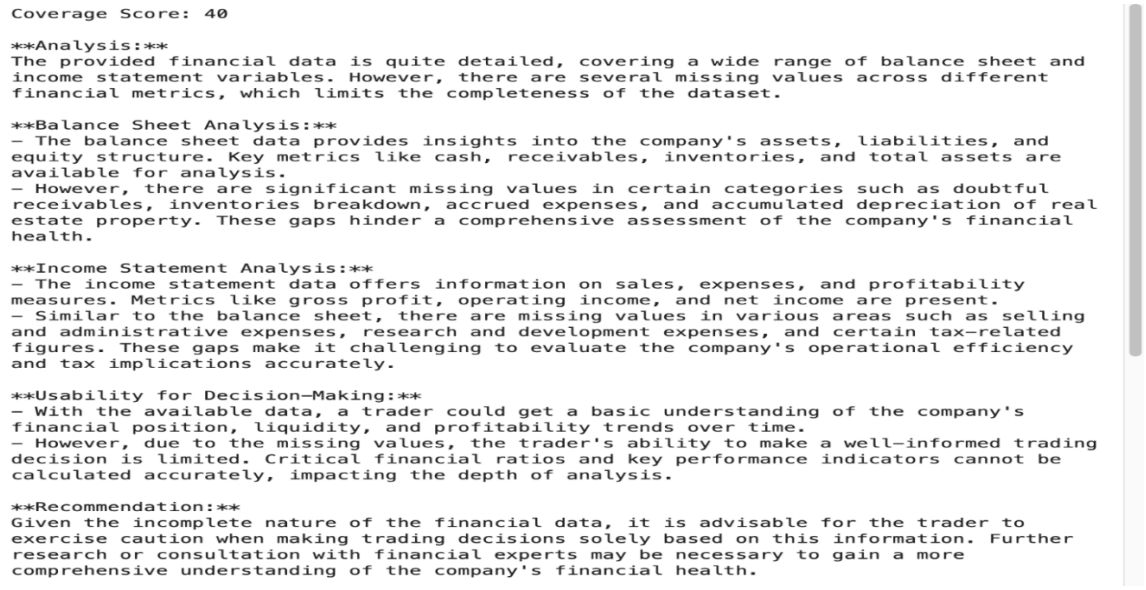
### Decision-Making:
- **Based on the data provided:**
  - The company has shown growth over the years, with increasing assets and revenues.
  - Profitability seems stable, with positive net incomes.
  - The company's capital structure and debt levels can be further analyzed once missing debt details are provided.

### Trading Decision:
- **Consider buying:** If the company's growth trajectory seems healthy and profitability is consistent, it might be
a good time to buy. However, a more thorough analysis of debt and specific expenses is needed for a more informed
decision.

By leveraging the available data and conducting further analysis into missing variables, you can assist the trader
in making a basic trading decision.
```

For example, Figure A3 above presents the actual model output for Sterling Infrastructure, Inc. (STRL), illustrating how the prompt-guided AI model produces a structured analysis and a numeric score. The model discusses STRL's liquidity, solvency, and capital structure, noting the data offers a comprehensive view of the company's financial position and performance. ChatGPT concludes that it might be a good time to buy; however, a more thorough analysis of debt and specific expenses is needed. The model ultimately assigns STRL a *CS* of 70, a moderately high score that reflects substantial but not perfect disclosure. As another example, Figure A4 below contains the model output for Siebert Financial Corp. (SIEB), a smaller firm with numerous null entries, which lead to a much lower score (40). ChatGPT explicitly advises that the trader should exercise caution in decision-making because of significant missing items. Based on these examples, we believe that the prompt design is suitable for our purposes.

Figure A4: Prompt-guided ChatGPT Output for Siebert Financial Corp. (SIEB)



```
Coverage Score: 40

**Analysis:**
The provided financial data is quite detailed, covering a wide range of balance sheet and
income statement variables. However, there are several missing values across different
financial metrics, which limits the completeness of the dataset.

**Balance Sheet Analysis:**
- The balance sheet data provides insights into the company's assets, liabilities, and
equity structure. Key metrics like cash, receivables, inventories, and total assets are
available for analysis.
- However, there are significant missing values in certain categories such as doubtful
receivables, inventories breakdown, accrued expenses, and accumulated depreciation of real
estate property. These gaps hinder a comprehensive assessment of the company's financial
health.

**Income Statement Analysis:**
- The income statement data offers information on sales, expenses, and profitability
measures. Metrics like gross profit, operating income, and net income are present.
- Similar to the balance sheet, there are missing values in various areas such as selling
and administrative expenses, research and development expenses, and certain tax-related
figures. These gaps make it challenging to evaluate the company's operational efficiency
and tax implications accurately.

**Usability for Decision-Making:**
- With the available data, a trader could get a basic understanding of the company's
financial position, liquidity, and profitability trends over time.
- However, due to the missing values, the trader's ability to make a well-informed trading
decision is limited. Critical financial ratios and key performance indicators cannot be
calculated accurately, impacting the depth of analysis.

**Recommendation:**
Given the incomplete nature of the financial data, it is advisable for the trader to
exercise caution when making trading decisions solely based on this information. Further
research or consultation with financial experts may be necessary to gain a more
comprehensive understanding of the company's financial health.
```

For the above task, we use OpenAI's GPT-3.5 Turbo as our model. It is the version of ChatGPT that was widely available to users from November 30, 2022 through 2023. That is, GPT-3.5 was the model that most retail investors had access to during our sample period (pre-event window: 11/30/2021-11/29/2022; and post-event window: 12/1/2022-11/20/2023). Note that in early 2023, GPT-4 was available only to a sub-set of users and required them to pay for subscription; and it became available to the general public (API developers) only around July 2023 (OpenAI, 2023). That is why we choose GPT-3.5 Turbo. Another reason is that employing the lighter model enables us to run the procedure efficiently (with low costs) at scale, covering thousands of firms. GPT-3.5 Turbo, albeit less advanced than GPT-4, is still a powerful model and, it was capable of differentiating firms based on the criteria specified in the prompt. In addition, GPT-4 was technically unable to perform the task at the level required, because its context (or input) length and token limits restricted its capacity to process the full scope of inputs. We believe that instead of using

an elite model beyond most users' reach, our model choice makes the empirical setup realistic, thereby capturing the effects of the publicly accessible AI. Future research may re-examine the method of deriving the *CS* measure with GPT-4 or GPT-5 to assess if this enhances the distinctions.

A.4. Multi-Temperature Sampling for Robustness

A challenge with LLM-based analyses is their potential variability – the same model may yield slightly different answers (and scores) when run multiple times, especially in non-deterministic settings. To address this issue and ensure our *CS* metric is more reliable, we use multi-temperature sampling and take the average* across the scores. This approach is consistent with the idea of using an ensemble of model outputs to smooth out randomness (Wang et al., 2023) and the suggestions by prior studies noting that LLM responses can vary with different sampling parameters. Such methodological robustness is particularly important, given the concerns about reproducibility and replicability in empirical research.

To be specific, for each firm we run eight separate inference passes in four different temperature settings: 0.0, 0.5, 0.7, and 1.0. A temperature is the parameter that controls the randomness/creativity of a model's output: temperature 0 makes the output fully deterministic (the model always picks the highest-probability next token), while higher temperatures introduce more randomness in responses. With temperature 1.0, for example, we get more varied phrasing and alternative interpretations of the data; at 0.0, we get the most likely, straightforward interpretation. We run the procedure twice with each of the four temperatures, yielding eight outputs and scores for each firm. We then compute the average of eight scores, effectively smoothing out model randomness. This can cancel out idiosyncratic fluctuations. If in one run the model's attention strayed or it slightly over-penalized a missing item, another run might counterbalance it. We find that this greatly improved consistency: the standard deviation of scores across firms is much lower with the averaged scores than with individual ones, as reported in Table 1.

This design choice reflects how generative AI tools are used in practice. Investors interact with LLMs through varied usage patterns, including differences in prompts, levels of exploration, and tolerance for narrative variation. Some users depend on deterministic summaries, while others participate in more exploratory interactions. The stochasticity inherent in LLM outputs therefore mirrors real-world diversity in user behavior rather than measurement noise alone. Our goal is not to achieve a perfectly deterministic firm-level score but to establish a stable classification of firms based on their exposure to AI-assisted interpretation. Assessing firms across multiple temperature settings captures the central tendency of how informative a firm's disclosures are across a realistic range of usage scenarios. Because treatment assignment relies on percentile cutoffs of the averaged score, residual variation across runs is unlikely to impact firm classification or the difference-in-differences analysis.

Moreover, averaging across temperatures mitigates any bias introduced by a particular setting. A very high temperature might produce an excessively optimistic or pessimistic narrative (due to a quirky wording or focus), whereas a low temperature might be overly conservative. The averaged coverage score (denoted

as *CSA*) reinforces the robustness of our metric. We essentially implement a form of self-consistency, where multiple "noisy" reasoning paths are synthesized into a final judgment (Wang et al., 2023).

Figure A5: Distributions of the Coverage Scores by Temperature

This figure shows the distribution of the Coverage Score (*CS*) for the eight cases, with two runs for each of the four temperature values (ranging from 0.0 to 1.0). Specifically, *CS1* and *CS2* are for the plots of *CS* at temperature 0.0; *CS3* and *CS4* are at 0.5; *CS5* and *CS6* are at 0.7; and *CS7* and *CS8* are at 1.0. The last one depicts the distribution of the score average (*CSA*) across all eight runs.

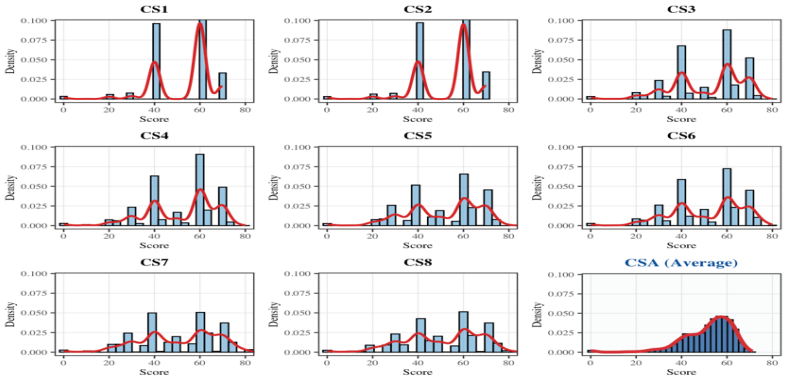


Figure A5 displays the distribution of scores for each run and the overall average (*CSA*) across all eight runs. Notably, the distributions are mainly bimodal. There is a higher-density cluster of firms with scores around the upper middle range (approximately 60-65) and another group in the lower range (around 40), with fewer companies in between and very few below 30 or above 90. This bimodality indicates two main groups: one with firms that have more comprehensive financial disclosures, enabling ChatGPT to assign higher scores, and another with more limited or fragmented disclosures, leading to moderate and generally lower scores. The bimodality offers a clearer picture—many firms fall into the categories of "quite coverable" versus "only partially coverable," with fewer in the middle. Practically, the information about bimodality in individual runs is helpful, as it offers a natural way to divide treatment and control groups. We base our cutoff on these distributions: firms in the top 30% of *CSA* are treated, while the rest serve as controls (that is, the 70th percentile of *CSA*, which is 58.40, as shown in Table 1).

As shown in the last plot of Figure A5, the average score (**CSA**) is quite smooth and nearly unimodal. Overall, this suggests that the *CSA* metric consistently reflects the underlying coverage and usefulness of firms' disclosures as evaluated by the LLM. We also manually reviewed a subset of model outputs alongside the corresponding firm disclosures, verifying that the model reliably follows the specified evaluation

INTERNET APPENDIX B: Robustness Tests

B.1. Placebo Tests with Pseudo Event Dates

B.1.1. Set-up and Choice of Dates

One concern in any difference-in-differences (DiD) analysis is whether the treatment effect genuinely reflects the intervention or instead captures spurious trends that would have emerged regardless of the timing of the intervention. Our empirical study exploits ChatGPT's release (on November 30, 2022) as an external shock, attributing post-release changes of our key measures in the treatment firms to ChatGPT's impact. However, if the treatment and control firms exhibited differential trends prior to its release, our DiD estimates could spuriously capture these pre-existing trends rather than the true effects.

Placebo tests can address this concern by re-estimating our specifications using "fake" event dates when no intervention occurred (Bertrand et al., 2004). If our results genuinely reflect ChatGPT's impact, we should not observe any consistent and significant effects after the placebo dates, given that treatment and control firms would have evolved similarly absent the actual intervention. Conversely, if placebo tests yield significant effects comparable to our main results, this would suggest our findings spuriously capture pre-existing differential trends rather than ChatGPT's impact, undermining the causal interpretation.

We implement placebo tests below by selecting three alternative event dates in 2021 before ChatGPT's actual release: (i) May 30, 2021; (ii) June 30, 2021; and (iii) November 30, 2021. The first two dates are chosen to include different quarters and macroeconomic conditions. The third date (November 30, 2021), which is exactly one year before its release, provides the most conservative test by holding constant seasonal effects and macroeconomic conditions that might systematically vary across calendar dates. This date is the latest possible placebo timing that can separate from our post-event window. We include a symmetric one-year window before and after the placebo date as in our main analyses. We then estimate the regression for each of our key dependent variables as in Eq. (6) in the main text. Given the fake dates, we replace the term *POST* with $POST^{PLB}$ in the placebo tests. We then estimate the equation for all of our key dependent variables. To save space, we provide only the results from the specification that includes both firm and date fixed effects. For the same reason, we do not report the marginal effects for the specifications that employ fractional logit estimation. Note in the result table that the three terms ($POST^{PLB}$, *TREAT*, and the intercept) are absorbed by including the two types of fixed effects in the specifications.

B.1.2. Placebo Test Results

Table B1 reports the results for ten key variables of interest with three pseudo event dates. In Panel A where May 30, 2021 is assumed as the event date, we find that the vast majority of treatment effects are statistically insignificant or have a wrong sign. For example, the columns that use the informed-trading measures (π^e, π^g, and π^b) as the dependent variable show that the coefficient on ($POST^{PLB}$ x *TREAT*) is positive and insignificant. As to the columns (for the price-impact and volatility measures) that use OLS

estimation, the coefficient is positive and significant (with a wrong sign again) when $Vola_{5m}$ or $Vola_{5m}^{mid}$ is used as the dependent variable. With the three order-imbalance measures, the sign of the coefficient is correct (positive), but none of the three cases is significantly associated with the outcome variable. This suggests that the treatment and control firms do not exhibit differential trends during periods when there is no intervention (release of ChatGPT).

Panel B where June 30, 2021 is used as the event date, the results are virtually the same. To make the alternative date distant from the first one used in Panel A, we choose November 30, 2021 as the fake event date in Panel C. By selecting this date (6 months away from the first date but still one full year earlier than the actual event date), we may be able to avoid any potential contamination (e.g., the anticipation effect). Interestingly, we find that the coefficient is significant at 5% or higher; however, this is again with the opposite sign, meaning that the interaction term is positively associated with the three order-imbalance measures (*ADOI, ADOI*^{R}, and *ADOI*^{NR}). These inconsistent results confirm that our main test results capture ChatGPT's impact rather than spuriously reflecting pre-existing trends.

The placebo test results also address several other issues regarding our findings. First, they rule out the possibility that our main results are driven by *selection effects* in which treatment firms systematically differ from control firms. If selection were driving the effects, the placebo tests should produce similar magnitudes and significance patterns, but they do not. Second, the results mitigate concerns about *anticipation effects*, where investors may have predicted ChatGPT's release and adjusted their behavior in advance. If anticipation were substantial, we would observe treatment effects emerging prior to the actual release date (November 30, 2022). However, the placebo test using November 30, 2021 (one year earlier) reveals no such anticipatory patterns for most outcome variables. Third, the sign reversals for the order-imbalance and volatility measures suggest that our main results reflect genuine *changes in trajectory* rather than simple continuations of pre-existing trends.

All in all, the above results corroborate that the treatment and control firms behaved similarly prior to ChatGPT's release. More importantly, our primary results are not driven by spurious pre-trends but capture the impact of ChatGPT on information asymmetry and trading activities in the treatment group of firms.

B.2. With an Alternative Method for Identifying the Treatment Group of Firms

To further check the robustness of our results, we use an alternative method to identify treatment firms. From the eight coverage scores (*CS1*-*CS8*) obtained for each firm, we discard the smallest and largest values to mitigate the potential influence of outliers. We then average the remaining six scores to construct *CSA*^{alt} for each firm. We next classify the sample firms in the top 30% of the *CSA*^{alt} distribution as the treatment group and others as the control group. The results of DiD analyses using this method are reported in Table B2 below. Note that the three terms (*POST*, *TREAT*, and the intercept) are absorbed by including

both firm and date fixed effects. When the dependent variable is bounded by [0, 1], fractional logit estimation is used, instead of OLS estimation. In that case, marginal effects are not reported to save space.

We find in Columns (1)-(3) of Table B2 that the effect of ChatGPT's release on informed-trading activities in the treatment firms is slightly weaker than that in the corresponding specifications reported in Tables 3. In Column (2), for example, the coefficient (-0.0746) on the interaction term (*POST* x *TREAT*) is smaller by about 25% in absolute terms than that of Column (4) in Table 3. However, the patterns are qualitatively similar. With the price-impact measures (λ^{GH} and λ^{S}) as the dependent variable, the magnitude of the coefficients on the interaction term shows a similar pattern. As to the result with the return volatility measures ($Vola_{5m}$ and $Vola_{5m}^{mid}$), the impact is equivalent to or even stronger than that of the corresponding specifications reported in Table 5. When the absolute order-imbalance measures (*ADOI, ADOI*^{R}, and *ADOI*^{NR}) are regressed in Columns (8)-(10), the size of the coefficients on (*POST* x *TREAT*) and their statistical significance are virtually the same as those observed in Table 6. Taken together, our main results are little affected by using the alternative method to identify the treatment group of firms.

Table B1
Results from Placebo Tests with Pseudo Event Dates

This table reports the results from placebo tests conducted by assuming that the event (the release of ChatGPT) occurred on different dates, instead of November 30 2022. Panels A-C contain the results obtained by assuming that the event occurred on May 30, 2021, June 30, 2021, and November 30, 2021, respectively; therefore, the pre- and post-event windows include a corresponding one-year interval before and after each of the three placebo event dates. The key variables are defined as follows. π^e: daily posterior probability of informed trading; π^g (π^b): daily posterior probability of informed buying (selling) on good (bad) news, estimated using daily buyer- and seller-initiated trades following Brennan, Huh, and Subrahmanyam (2018); λ^{GH} and λ^S: price-impact parameter (or the adverse-selection component of trading costs) estimated based on Glosten and Harris (1988) and Sadka (2006), respectively, using intraday dollar order flows within each day (given the extremely large kurtoses and small values, the two measures are winsorized at the 0.5th and 99.5th percentiles, and also multiplied by 10^4); $Vola_{5m}$: daily standard deviation of intraday returns (using transaction prices) sampled at a five-minute frequency within each trading day; $Vola_{5m}^{mid}$: daily standard deviation of intraday returns (using bid-ask quote midpoints) sampled at a five-minute frequency within each trading day; *ADOI*: absolute dollar order imbalance computed as $\frac{|dbuy-dsell|}{dbuy+dsell}$, where *dbuy* and *dsell* are daily dollar value of shares bought and that of sold, respectively; $ADOI^R$($ADOI^{NR}$): absolute dollar order imbalance caused by retail (non-retail or institutional) investors, for which trades by retail investors are identified following the algorithm proposed by Barber et al. (2024) (to obtain daily buys/sells and quote midpoints, each trade is matched with quotes based on the Holden-Jacobsen (2014) algorithm); $POST^{PLB}$: one if the observation occurs after the placebo event date, and zero otherwise; *TREAT*: one for firms in the top 30% of the *CSA* distribution, and zero otherwise; and ($POST^{PLB}$ x *TREAT*): an interaction term between the two indicator variables. Note that the three variables ($POST^{PLB}$, *TREAT*, and the intercept) are absorbed by including both firm and date fixed effects. Three firm characteristics as control variables are obtained as below. *SIZE*: natural log of market capitalization; *BTM*: ratio of book equity to market equity, for which the previous quarter-end book value is matched with daily market-equity values within each quarter; and *VOLA*: daily standard deviation (multiplied by 10^2) of intraday returns from WRDS. When the dependent variable is bounded by [0, 1], fractional logit estimation is used (marginal effects are not reported for brevity), instead of OLS estimation. Standard errors are clustered by firm and date, and *t*-statistics are reported in parentheses. The coefficients significantly different from zero at the 10%, 5%, and 1% levels are indicated by *, **, and ***, respectively.

	Placebo Tests									
	Panel A: With a Placebo Event Date on May 30, 2021									
	Fractional Logit			OLS				Fractional Logit		
Variables \ Dep. Var.	π^e	π^g	π^b	λ^{GH}	λ^S	$Vola_{5m}$	$Vola_{5m}^{mid}$	*ADOI*	$ADOI^R$	$ADOI^{NR}$
$POST^{PLB}$ x TREAT	0.0476	0.0339	0.0480	0.0183	0.0273	0.0003***	0.0002***	0.0127	0.0030	0.0193
	(1.29)	(0.94)	(1.25)	(1.02)	(0.78)	(6.15)	(5.29)	(0.82)	(0.19)	(1.25)
SIZE	0.3129***	0.4116***	0.0278	-0.0466***	0.0064	-0.0005***	0.0001	-0.2746***	-0.2544***	-0.3106***
	(8.84)	(11.96)	(0.78)	(-3.85)	(0.20)	(-6.82)	(0.12)	(-17.37)	(-15.23)	(-19.06)
BTM	-0.0001	0.0001	-0.0011	-0.0000	-0.0000	0.0000	0.0000	-0.0004	-0.0000	-0.0002**
	(-0.33)	(0.28)	(-0.23)	(-0.35)	(-0.59)	(0.76)	(0.69)	(-1.33)	(-1.34)	(-2.00)
VOLA	0.0011	0.0112	-21.2400***	-0.0176**	-0.0034			-0.0059	7.0540***	-0.0028
	(0.09)	(1.81)	(-5.24)	(-2.48)	(-0.40)			(-1.09)	(4.20)	(-0.52)
Firm FE	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Date FE	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Clustering	Firm & Date	Firm & Date	Firm & Date	Firm & Date	Firm & Date	Firm & Date	Firm & Date	Firm & Date	Firm & Date	Firm & Date
Obs	705,145	705,259	704,797	1,034,291	1,105,920	1,109,652	1,109,198	1,109,500	1,080,191	1,088,553
Pseudo/Adjusted R^2	0.0987	0.0536	0.0559	0.0088	0.0045	0.5677	0.4124	0.3953	0.4451	0.4040

(Table B1: continued)

	Panel B: With a Placebo Event Date on June 30, 2021									
	Fractional Logit			OLS				Fractional Logit		
Variables \ Dep. Var.	π^e	π^g	π^b	λ^{GH}	λ^S	$Vola_{5m}$	$Vola_{5m}^{mid}$	*ADOI*	*ADOI*^{R}	*ADOI*^{NR}
POST^{PLB} x TREAT	0.0515	0.0365	0.0480	0.0219	0.0481	0.0002***	0.0002***	0.0017	-0.0081	0.0082
	(1.42)	(1.03)	(1.28)	(1.16)	(1.47)	(4.22)	(5.45)	(0.11)	(-0.50)	(0.53)
SIZE	0.3056***	0.4093***	0.0075	-0.0475***	0.0024	-0.0006***	-0.0001	-0.2617***	-0.2459***	-0.3004***
	(9.07)	(12.26)	(0.21)	(-3.83)	(0.08)	(-8.09)	(-1.39)	(-17.21)	(-15.17)	(-18.77)
BTM	-0.0002	0.0001	-0.0146	-0.0000	-0.0000	0.0000	0.0000	-0.0004	-0.0000	-0.0003**
	(-0.67)	(0.24)	(-0.47)	(-0.41)	(-0.84)	(0.83)	(0.73)	(-1.33)	(-1.37)	(-3.00)
VOLA	0.0006	0.0110*	-25.3700***	-0.0146***	-0.0027			-0.0058	6.7640***	-0.0028
	(0.05)	(1.77)	(-5.64)	(-2.86)	(-0.29)			(-1.05)	(4.04)	(-0.52)
Firm FE	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Date FE	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Clustering	Firm & Date	Firm & Date	Firm & Date	Firm & Date	Firm & Date	Firm & Date	Firm & Date	Firm & Date	Firm & Date	Firm & Date
Obs	715,490	715,644	715,184	1,122,655	1,119,183	1,122,843	1,122,384	1,122,678	1,092,137	1,100,346
Pseudo/Adjusted R^2	0.1028	0.0561	0.0579	0.0081	0.0043	0.5743	0.4182	0.3949	0.4444	0.4283
	Panel C: With a Placebo Event Date on November 30, 2021									
	Fractional Logit			OLS				Fractional Logit		
Variables \ Dep. Var.	π^e	π^g	π^b	λ^{GH}	λ^S	$Vola_{5m}$	$Vola_{5m}^{mid}$	*ADOI*	*ADOI*^{R}	*ADOI*^{NR}
POST^{PLB} x TREAT	0.0052	-0.0369	0.0357	0.0082	0.0259	0.00001	0.0001***	-0.0455***	-0.0415**	-0.0425**
	(0.16)	(-1.08)	(1.03)	(0.43)	(0.65)	(0.20)	(2.64)	(-2.60)	(2.39)	(-2.38)
SIZE	0.3080***	0.4871***	-0.0386	-0.0570***	-0.0000	-0.0012***	-0.0004***	-0.2162***	-0.2115***	-0.2696***
	(10.10)	(14.99)	(-1.18)	(-4.22)	(-0.001)	(-15.53)	(-6.97)	(-17.15)	(-14.29)	(-18.60)
BTM	0.0733**	0.1750***	-0.0460	-0.0131	0.0250	0.0003***	0.0000	0.0498***	0.0306	0.0423**
	(2.42)	(5.03)	(-1.58)	(-0.77)	(0.42)	(4.30)	(0.88)	(3.41)	(1.54)	(2.20)
VOLA	-0.0011	-0.0369	-22.0300***	-0.0147**	0.0049			-0.0040	6.1730***	-0.0020
	(-0.06)	(-1.08)	(-6.79)	(-2.13)	(0.40)			(-0.63)	(4.08)	(-0.33)
Firm FE	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Date FE	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Clustering	Firm & Date	Firm & Date	Firm & Date	Firm & Date	Firm & Date	Firm & Date	Firm & Date	Firm & Date	Firm & Date	Firm & Date
Obs	744,457	744,602	743,972	1,172,012	1,169,006	1,172,075	1,171,566	1,172,025	1,130,969	1,139,054
Pseudo/Adjusted R^2	0.1012	0.0586	0.0553	0.0075	0.0028	0.6000	0.4348	0.3996	0.4443	0.4131

Table B2
Results with an Alternative Method for Identifying the Treatment Group of Firms

This table reports the results of DiD analyses using an alternative method to identify treatment firms. The two-year sample period includes a one-year window before and after the release of ChatGPT on November 30, 2022, with the pre-event window spanning from 11/30/2021 to 11/29/2022, and the post-event window from 12/01/2022 to 11/29/2023. From the eight coverage scores (*CS1*-*CS8*) obtained for each firm, the smallest and largest values are discarded to mitigate the potential influence of outliers. We then average the remaining six scores to construct CSA^{alt}. Firms in the top 30% of the CSA^{alt} distribution are classified as the treatment group, and all others serve as the control group. Most of the key variables are defined in Appendix B1. The remaining variables are defined as follows. *POST*: one if the observation occurs after the event date, and zero otherwise; *TREAT*: one for firms in the top 30% of the CSA^{alt} distribution, and zero otherwise; and (*POST* x *TREAT*): an interaction term between the two indicator variables. Note that the three variables (*POST*, *TREAT*, and Intercept) are absorbed by including both firm and date fixed effects. Three firm characteristics as control variables are obtained as below. *SIZE*: natural log of market capitalization; *BTM*: ratio of book equity to market equity, for which the previous quarter-end book value is matched with daily market-equity values within each quarter; and *VOLA*: daily standard deviation (multiplied by 10^2) of intraday returns from WRDS. When the dependent variable is bounded by [0, 1], fractional logit estimation is used (marginal effects are not reported for brevity), instead of OLS estimation. Standard errors are clustered by firm and date, and *t*-statistics are reported in parentheses. The coefficients significantly different from zero at the 10%, 5%, and 1% levels are indicated by *, **, and ***, respectively.

Effect of ChatGPT: With an Alternative Method for Identifying Treatment Firms

Estimation Method:	Fractional Logit			OLS				Fractional Logit		
Dep. Var.:	π^e	π^g	π^b	λ^{GH}	λ^S	$Vola_{5m}$	$Vola_{5m}^{mid}$	*ADOI*	$ADOI^R$	$ADOI^{NR}$
Variables	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)
POST x TREAT	-0.0597*	-0.0746**	-0.0201	-0.0469*	-0.1015*	-0.0003***	-0.0002***	0.0324**	0.0472***	0.0456***
	(-1.77)	(-2.09)	(-0.59)	(-1.72)	(-1.88)	(-5.01)	(-5.52)	(2.40)	(3.21)	(3.08)
SIZE	0.1845***	0.3557***	-0.0713**	-0.0728***	-0.0776	-0.0020***	-0.0009***	-0.1833***	-0.1710***	-0.2523***
	(4.53)	(12.10)	(-2.43)	(-3.75)	(-1.59)	(-27.55)	(-19.03)	(-15.02)	(-11.32)	(-16.71)
BTM	0.0581**	0.1481***	-0.0436*	0.0237	0.1318	0.0002***	-0.0001*	0.0492***	0.0677***	0.0536***
	(2.39)	(5.72)	(-1.82)	(1.16)	(1.51)	(3.28)	(-1.93)	(5.53)	(5.94)	(4.25)
VOLA	-1.0830	0.0033	-13.3400***	0.0096	0.0306			0.0008	5.6270***	0.0003
	(-0.16)	(0.31)	(-9.10)	(0.92)	(1.29)			(0.08)	(6.73)	(0.04)
Firm FE	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Date FE	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Clustering	Firm & Date	Firm & Date	Firm & Date	Firm & Date	Firm & Date	Firm & Date	Firm & Date	Firm & Date	Firm & Date	Firm & Date
Obs	784,743	784,781	784,189	1,211,118	1,208,131	1,210,971	1,210,234	1,211,134	1,125,605	1,136,000
Pseudo/Adjusted R^2	0.0865	0.0511	0.0462	0.0109	0.0053	0.6378	0.4315	0.4093	0.4545	0.4317