

CAN THE CROWD FORECAST THE FIRM?

**COMPANY-LEVEL REVIEW SENTIMENT AND FIRM
PERFORMANCE**

**ELLINOR LINDÉ
EBBA NYSTRÖM**

Bachelor Thesis

Stockholm School of Economics

2026

Can the Crowd Forecast the Firm? Company-Level Review Sentiment and Firm Performance

Abstract:

We examine if sentiment embedded in company-level customer reviews contains information relevant to stock returns and firm fundamentals. Using web scraped Trustpilot reviews for Nordic listed retail firms from 2014 to 2025, we construct a text-based measure of abnormal customer sentiment and test whether it predicts subsequent excess returns and revenue growth. The strongest evidence appears for firm fundamentals, where abnormal customer sentiment positively and significantly predicts subsequent revenue growth. The results provide limited evidence that abnormal sentiment predicts short-term excess returns, while the coefficients are directionally consistent with prior product-level findings, they are not statistically significant. Furthermore, negative abnormal sentiment appears more return-relevant than positive abnormal sentiment when customer opinions are more dispersed. Overall, company-level reviews contain value-relevant information, but this information is reflected more clearly in future revenue growth than in short-term stock price movements.

Keywords: Company-Level Reviews, Trustpilot, Customer Review Sentiment, Abnormal Sentiment, Excess Returns, Revenue Growth

Authors: Ellinor Lindé (50936), Ebba Nyström (50906)

Supervisor: Alexander Mafael, Head of Research, Center for Retailing

Examiner: Lily Gao

Acknowledgement: We extend our sincere gratitude to our supervisor, Alexander Mafael, for his continuous support. We are grateful for his willingness to answer our many questions and for serving as a valuable sounding board for our ideas and methodological choices. His input and guidance have been greatly valuable and appreciated.

Bachelor Thesis

Bachelor Program in Retail Management

Stockholm School of Economics

© Ellinor Lindé and Ebba Nyström, 2026

1. Introduction

The rise of e-commerce and digital review platforms has changed the way consumers evaluate firms. Online review platforms aggregate individual customer experiences into publicly observable information that extends beyond traditional financial disclosures (Huang, 2018). Consumers rely heavily on reviews when making purchase decisions. Raketa (2024) and Qiu and Zhang (2024) document that 98% of consumers use online reviews to inform purchase decisions, and are more likely to purchase products with positive reviews. Prior literature confirms that this behavior translates into tangible firm outcomes, with online reviews influencing both consumer purchasing behavior and firm sales (Chevalier and Mayzlin, 2006; Liu, 2006; Ludwig et al., 2013). Since firms' future cash flows depend on their ability to create value for customers, information embedded in customer reviews may thus be relevant not only to consumers but also to the firm and its investors.

Consequently, the financial relevance of customer opinions has received growing attention in academic research. Prior studies show that product review sentiments predict future demand, sales, and stock market outcomes, suggesting that aggregated customer opinions may contain forward-looking information about firm performance (Archak et al., 2011; Wu et al., 2020; Jin, 2023). However, this evidence focuses exclusively on product-level reviews, which evaluate specific goods on platforms such as Amazon and JD.com. It remains unclear whether this evidence generalizes to company-level reviews, where customers evaluate the overall firm experience rather than a single product. In this setting, the unit of evaluation shifts from a specific product to the firm as a whole. Company-level reviews may capture a broader set of customer experiences, including delivery, returns, customer support, reliability, and trust, and therefore provide a more holistic signal of how customers experience the firm (Parasuraman et al., 1988; Richard and Allaway, 1993). Accordingly, whether sentiment embedded in broader firm-level evaluations contains information relevant to equity markets and future firm fundamentals remains an open question.

The purpose of this thesis is therefore to address this gap in the existing literature by examining whether abnormal sentiment derived from company-level customer reviews predicts subsequent firm performance. In this thesis, firm performance is used as an

umbrella term for two distinct outcomes. First, we examine market-based performance, measured by subsequent excess returns. Second, we examine firm fundamentals, proxied by subsequent revenue growth. The first outcome captures whether customer sentiment is reflected in stock prices, while the second captures whether customer sentiment contains information about underlying demand and operating performance. More specifically, we analyze the following three research questions:

1. Does abnormal sentiment from company-level customer reviews predict subsequent excess returns?
2. Does negative abnormal sentiment from company-level customer reviews have stronger return relevance than positive abnormal customer sentiment?
3. Does company-level abnormal customer review sentiment predict subsequent revenue growth?

The answer to these questions is theoretically ambiguous. Under the efficient market hypothesis, publicly available information should be incorporated into stock prices quickly (Fama, 1970). If investors fully observe and process customer reviews, review sentiment should not predict future excess returns. However, company-level reviews may still contain value-relevant information if they aggregate dispersed customer experiences that investors do not fully process or translate into expectations about future demand. In this sense, reviews can be interpreted as a form of serendipitous information, a concept introduced by Subrahmanyam and Titman (1999), in which many individual customer experiences collectively provide information about firm value. At the same time, company-level reviews are subjective, noisy, and unevenly distributed across firms, which may limit their usefulness as an investment signal.

To address these questions, we construct a firm-level text-based sentiment measure from Trustpilot, a company-level review platform that has received limited attention in the financial markets context. The dataset is created through web scraping and covers a sample of Nordic listed retail firms over the period 2014 to 2025. The analysis focuses on abnormal sentiment, capturing deviations in customer evaluations from firms' recent historical sentiment levels, and examines whether these deviations predict subsequent excess returns and revenue growth in both Fama-MacBeth and fixed-effects panel regressions.

Our analysis uncovers several new findings. First, we find strong evidence that abnormal customer sentiment is positively and significantly associated with subsequent revenue growth. For stock returns, our analysis provides limited support that abnormal company-level review sentiment predicts short-term excess returns overall, although the coefficients are generally positive. Furthermore, we find that negative sentiment may be more informative than positive sentiment when customer opinions are more dispersed. Taken together, the findings suggest that company-level review sentiment may be more informative about future demand-related fundamentals than about short-term stock price movements.

Our thesis contributes to the literature in two distinct ways. First, it extends the review literature from a product-level to a company-level setting by examining whether firm-level customer evaluations contain information relevant to both capital market outcomes and firm fundamentals. This extension is important because it tests whether the financial relevance of reviews generalizes from product-specific settings to broader firm-level customer evaluations. Second, we introduce a new firm-level, text-based sentiment measure constructed from Trustpilot reviews, thereby contributing a new data source and measurement approach to a literature that has predominantly relied on product-level review platforms.

The remainder of the paper is structured as follows. Section 2 reviews the related literature, positions the thesis relative to prior research, and discusses the three hypotheses in greater detail. Section 3 describes the data collection, sentiment measurement, variable construction, and sample characteristics. Section 4 outlines the empirical methodology and presents the main results. Section 5 presents the robustness tests. Finally, Section 6 discusses the findings in relation to prior literature, states implications for academia, managers, and investors, and addresses limitations and future research.

2. Theoretical Background

This section develops the theoretical foundation for examining whether company-level customer review sentiment contains information relevant to stock returns and firm fundamentals. It begins by reviewing prior research on how customer opinions can contain information about firm value and performance. It then positions the thesis relative to

existing literature and develops the hypotheses tested in the empirical analysis.

2.1 Literature Review

Customer opinions may constitute a non-traditional source of information about firm value. Subrahmanyam and Titman's (1999) concept of serendipitous information provides a theoretical foundation for this view. Unlike traditional financial information, serendipitous information captures value-relevant signals that emerge incidentally from individuals' everyday observations, such as consumer experiences, product usage, and perceptions of demand conditions. Although individual observations may be noisy, their aggregation can reveal information about a firm's underlying value. Online reviews can be interpreted as modern representation of this concept, as they consolidate dispersed customer experiences into observable public sentiment. Furthermore, Subrahmanyam and Titman (1999) suggest that serendipitous information is likely to be more prevalent in industries characterized by broad customer bases and high demand uncertainty.

The idea that dispersed opinions can be informative also relates to the wisdom of crowds concept. Surowiecki (2004) argues that groups can produce accurate collective judgments that can exceed those of individuals. Consistent with this view, Chen et al. (2014) show that the tone of investor-generated articles and comments on Seeking Alpha predicts both future stock returns and earnings surprises. Their findings support the wisdom of crowds hypothesis by suggesting that aggregated opinions from dispersed investors contain value-relevant information not yet fully reflected in analyst forecasts or market prices, rather than merely reflecting noise or investor hype. Taken together, serendipitous information and the wisdom of crowds concept provide a theoretical foundation for the view that consumer opinions, when aggregated, may capture firm-level information relevant to subsequent financial market outcomes.

This theoretical view is supported by empirical research on customer satisfaction, particularly studies using the American Customer Satisfaction Index (ACSI). Fornell et al. (2016) find that customer satisfaction predicts future earnings surprises, and Ittner et al. (2009) show that ACSI scores predict future operating income. Some studies also reveal that firms with higher and improving customer satisfaction scores achieve stronger financial returns (Aksoy et al., 2008; Fornell et al., 2006; Fornell et al., 2016). However, the evidence is not unambiguous. Other studies find limited or no evidence of abnormal

long-run stock returns in response to satisfaction (Bell et al., 2014; Jacobson and Mizik, 2009; Ittner et al., 2009). Taken together, while the evidence on stock return predictability remains mixed, this literature suggests that customer satisfaction contains information about firm fundamentals.

While customer satisfaction measures provide static indicators of customer perceptions, related literature examines how these perceptions are communicated through word-of-mouth (WOM). Hennig-Thurau et al. (2002) show that WOM directly reflects underlying customer satisfaction and commitment. Chen and Xie (2008) relate this perspective to the online context by conceptualizing online consumer reviews as a modern digital form of WOM based on user experiences. This suggests that online reviews serve as a timely and observable reflection of customer satisfaction, making them relevant for studying whether customer opinions are reflected in financial markets.

Building on the view of online reviews as digital WOM, prior studies examine whether review-based signals are reflected in stock market outcomes. One strand of this literature focuses on numerical ratings. Luo et al. (2013) demonstrate that social media is a leading indicator of firm equity value, where consumer ratings have the highest predictive power for firm returns and risks. Additionally, Huang (2018) finds that customer ratings on Amazon.com contain information about stock prices, where abnormal customer ratings positively impact subsequent stock returns, particularly when limits to investor attention are more binding. Taken together, these studies suggest that customer ratings can aggregate dispersed customer information that investors do not immediately incorporate into stock prices.

Alongside numerical ratings, textual review content has also been examined for its informational value in explaining stock market outcomes. Archak et al. (2011) show that textual content in product reviews has a predictive power for consumer behavior, explaining variation in product demand above the valence and volume of reviews. Wu et al. (2020) extends this argument to capital markets and find that textual product reviews on JD.com predict long-term firm-level stock returns. Further, Jin (2023) shows that review tone on Amazon.com predicts a firm's abnormal stock returns and risk, where the impact on stock prices is channeled through its effects on earnings. Overall, this literature indicates that textual review content captures dimensions of customer evaluations beyond numerical ratings that can help explain stock market outcomes.

Beyond sentiment itself, the literature also points to review volume as a financially relevant signal. Luo and Zhang (2013) establish that customer buzz is related to firm value, while Tirunillai and Tellis (2012) find that the volume of online chatter has strong predictive power for stock returns and trading volume. Similarly, Babić Rosario et al. (2016) show that the volume of electronic-WOM (eWOM) is more strongly associated with sales than valence, suggesting that the sheer quantity of consumer opinions carries information beyond their average tone. The authors further find that eWOM variability negatively affects sales, indicating that disagreement among consumers may increase uncertainty about firm or product quality, increasing functional risk. This suggests that the quantity and distribution of consumer opinions expressed about a firm carry financial relevance alongside the sentiment those opinions convey.

Whilst the preceding literature established that both sentiment and volume of consumer reviews carry financial relevance, a related strand of research has examined whether positive and negative sentiment contribute equally to this effect. Prospect theory provides a theoretical basis for the asymmetric effect of negative and positive sentiment. Kahneman and Tversky (1979) argue that individuals evaluate losses more strongly than gains of equal magnitude. This principle, referred to as loss aversion, implies that negative information receives greater psychological weight than positive information. In the context of customer reviews, negative sentiment may therefore be more salient than equally positive sentiment. Furthermore, Anderson (1998) shows that negative customer experiences generate stronger WOM than equally positive experiences, suggesting that this asymmetry is also reflected in how consumers express their opinions.

Empirical evidence from product review settings supports this asymmetric view. Wu et al. (2025) establish that while consumer reviews broadly predict future stock return, it is specifically negative customer sentiment that drives this effect, with positive sentiment contributing comparatively little to return predictability. Supporting this, Tirunillai and Tellis (2012) find that negative product review chatter has a stronger relationship with stock returns and trading volume than positive review chatter. Together, these studies provide direct empirical support for the asymmetry predicted by prospect theory in a review-based financial markets context.

Beyond stock returns, prior literature examines whether consumer sentiment predicts firm fundamentals. This literature suggests that the relationship between consumer

opinions and financial outcomes operates through underlying economic performance. At the product-level, review sentiment and volume affect sales outcomes across several contexts, including book sales, box office revenue, and product conversion rates (Chevalier and Mayzlin, 2006; Liu, 2006; Ludwig et al., 2013). These studies suggest that consumer opinions translate into observable changes in demand.

Extending this logic to the firm-level, Huang (2018) finds that abnormal customer ratings predict not only subsequent stock returns, but also revenue and earnings surprises. This suggests that improved customer evaluations can precede improvements in firm performance. Consistent with this argument, Wu et al. (2025) show that negative review sentiment predicts future revenue, earnings surprises, and cash flow shocks, suggesting that customer opinions contain forward-looking information about firms' fundamentals. Jin (2023) further shows that review tone predicts forthcoming quarterly sales and earnings surprises, and that the effect of review tone on stock prices operates in part through its relation to future earnings. Taken together, this literature suggests that review sentiment carries predictive content for firm-level economic outcomes beyond stock returns.

2.2 Theoretical Contribution and Positioning

The reviewed literature establishes that customer review sentiment contains predictive information about stock returns and firm fundamentals, however this evidence is based primarily on product-level reviews. Considerably less attention has been given to company-level reviews as a distinct source of customer sentiment. To our knowledge, Cho et al. (2024) provide the most direct empirical precedent, finding that sentiment extracted from online hotel reviews predicts future booking demand. Although the study is situated in the hospitality industry, its findings support the argument that customer sentiment from company-level reviews can shape downstream economic outcomes by altering perceptions of firm quality and future demand.

Company-level reviews may capture a broader signal of customer experiences than product reviews, as customer evaluations often extend beyond the core product itself. Parasuraman et al. (1988) show that consumers assess service quality across five dimensions, tangibles, reliability, responsiveness, assurance, and empathy, while Richard and Allaway (1993) show that customer choice reflects both outcomes and process quality. At the same time, Zhu and Zhang (2010) show that the impact of online product reviews varies across

product types and consumer contexts, suggesting that evidence from product reviews can be highly context-specific. Taken together, this literature suggests that company-level reviews may capture customer experiences that extend beyond evaluations of individual products.

The empirical setting provides an additional contribution. Yu et al. (2013) demonstrate that different media channels have heterogeneous effects on stock performance, suggesting that the predictive relevance of sentiment may be platform dependent. Whilst prior studies have relied predominantly on platforms such as Amazon or JD.com, this thesis constructs a text-based sentiment measure from Trustpilot. Trustpilot is particularly suitable because it provides company-level customer reviews rather than product-specific evaluations, capturing broader customer experiences, including delivery, returns, customer support, reliability, and trust. Prior research using Trustpilot reviews in e-commerce settings shows that such reviews capture several stages of the customer journey and multiple sources of customer satisfaction and dissatisfaction (Davoodi et al., 2026). Unlike social media platforms such as Twitter/X, Trustpilot contains explicit customer reviews rather than a broader mix of consumer, investor, media, and social commentary. It also differs from location-oriented review platforms, such as Google Reviews, where reviews are often attached to individual stores rather than the company as a whole. Trustpilot, therefore, provides a more suitable basis for constructing consolidated firm-level sentiment measures. This allows the thesis to examine whether company-level review sentiment has financial relevance in a platform setting that has received limited attention in financial market research.

Taken together, this thesis makes two primary contributions. First, it examines whether sentiment derived from company-level customer reviews predicts subsequent stock returns and revenue growth, extending on literature that has focused predominantly on product-level reviews. Second, it constructs a firm-level, text-based sentiment measure from Trustpilot as a previously unexplored review platform.

2.3 Hypotheses

Building on this literature, we develop three hypotheses that extend prior review-based findings to a company-level setting by testing whether abnormal sentiment in firm-level customer reviews predict subsequent excess returns and revenue growth.

Customer reviews aggregate dispersed customer experience and may convey information about the firm that is not yet reflected in stock prices. Prior product-level research suggests that this information is positively related to stock market outcomes, as higher customer ratings and more favorable review sentiment predicts stronger subsequent stock returns (Luo et al., 2013; Huang, 2018; Wu et al., 2020; Jin, 2023). Extending this logic to company-level reviews, higher abnormal sentiment may indicate that customers' assessment of a firm has improved relative to its recent history. Such improvements should represent favorable information about future demand and firm value. If investors do not immediately incorporate this information into prices, firms with higher abnormal customer sentiment should earn higher subsequent excess returns. Therefore, we hypothesize:

H1: *Firms with higher abnormal customer review sentiment experience higher subsequent excess returns.*

Prospect theory suggests that negative information tends to receive greater attention and weight in decision-making (Kahneman and Tversky, 1979). Furthermore, prior literature suggests that negative customer sentiment in product reviews has a greater effect on stock market outcomes than positive sentiment (Wu et al., 2025; Tirunillai and Tellis, 2012). Extending this logic to company-level reviews, negative abnormal sentiment may attract greater investor attention than positive abnormal sentiment since unfavorable customer experiences are likely to be perceived as more informative. Negative abnormal sentiment should thus exhibit a stronger association with subsequent excess returns than positive abnormal sentiment. Therefore, we hypothesize:

H2: *Negative abnormal customer review sentiment exhibits a stronger association with subsequent excess returns than positive abnormal customer review sentiment.*

Previous literature further demonstrates that online reviews and customer sentiment are associated with future sales and operating performance. Empirical evidence suggests that review sentiment predicts product demand, conversion rates, sales, and earnings surprises (Chevalier and Mayzlin, 2006; Liu, 2006; Ludwig et al., 2013; Huang, 2018; Wu et al., 2025; Jin, 2023). Extending this evidence to company-level reviews, abnormal customer sentiment may capture information about changes in customer experiences and demand conditions at the firm-level. If abnormal sentiment reflects shifts in customer

experience, it should also be associated with subsequent changes in firm-level demand. Therefore, we hypothesize:

H3: *Firms with higher abnormal customer review sentiment experience higher future revenue growth.*

Together, the hypotheses test whether company-level customer sentiment contains information relevant to both market-based outcomes and firm fundamentals.

3. Data

The following section describes the data and variable construction used to examine whether company-level review sentiment contains information relevant to subsequent stock returns and revenue growth. We begin by outlining the data collection process, sample construction, and sentiment measurement procedure. Next, we define the dependent, explanatory, and control variables used in the empirical analysis. Finally, we present descriptive statistics and discuss data treatment applied prior to the analysis.

3.1 Data Collection and Sample Construction

We compile a sample of consumer discretionary and consumer staple firms listed on the Nordic stock exchanges: OM, OB, HLSE, CPSE, and ICSE¹. These industries are selected because they primarily serve end consumers, which makes direct consumer interaction and, consequently, consumer reviews more likely. We retrieve firm-level monthly returns, stock exchanges, and quarterly accounting data from Capital IQ and monthly Fama-French factors from the Kenneth French Data Library.

Following Huang (2018), we match monthly stock returns with accounting variables using a quarterly lag. Specifically, returns in a given month are matched with accounting information from the most recent prior fiscal quarter. This lag reduces look-ahead bias by ensuring that the accounting variables used in the regressions reflect information that investors could have observed at the time of return realization.

We collect customer review data from Trustpilot for the sample firms. Trustpilot is an online platform where customers provide feedback on companies based on both online and

¹ Exchange codes follow Capital IQ notation.

offline experiences. Unlike product-specific platforms, Trustpilot captures a broad range of interactions, including in-store service, delivery, and overall brand experience. Further, it is an international platform where users across the world leave reviews on companies, not only Nordic users. This makes it suitable for measuring firm-level sentiment as it captures global customer opinions.

Through web scraping, we collect review-level data for each firm and its associated brands. We map each parent firm to its corresponding Trustpilot websites through company disclosures and public information as firms differ in their organizational structure, frequently operating multiple brands and country-specific websites. As a result, a single listed firm may be associated with several Trustpilot links. All relevant links are then included in the scraping process to ensure that the firm-level sentiment measure captures the full set of customer reviews. For each review, we extract the review title, review text, star rating, and timestamp. The sample is restricted to firms within the financial portfolio that have a presence on Trustpilot. Firms with a limited number or no observable reviews are excluded. This matching procedure between the financial portfolio and review data results in a final sample of 81 firms over the period first of January 2014 to the last of December 2025, resulting in 2,116,158 scraped reviews.

The initial scraping procedure captures all available reviews for the Trustpilot links associated with each parent firm. However, some firm’s brand portfolios include acquired brands with reviews posted before the acquisition. To account for this, we map acquisition dates for each acquired brand within the firm portfolio. Reviews associated with an acquired brand are included from the announced acquisition date, provided that the transaction gives the parent firm majority ownership. This approach ensures that the sentiment measures are based on brands that can be reasonably linked to the parent firm’s operations. Since announcement dates may differ from legal closing or full operational integration, the timing rule may introduce some measurement noise around acquisitions. However, it provides a consistent and observable criterion for assigning brand-level reviews to parent firms. This procedure reduces the number of reviews to 1,739,568.

3.2 Sentiment Measurement and Aggregation

To measure the tone of each review, this thesis applies the pretrained multilingual sentiment model `cardiffnlp/twitter-xlm-roberta-base-sentiment`, developed by Cardiff NLP.

The model is selected due to its capability of processing text across multiple languages and is pretrained on approximately 198 million social media texts (Barbieri et al., 2022). This feature is essential given the international scope of the Trustpilot reviews, which contain customer feedback in several languages. Many sentiment analysis approaches are developed primarily for English text or require translation before classification, which would either reduce the sample or add a computationally costly preprocessing step that may introduce translation-based measurement error. The multilingual structure of the model therefore allows the analysis to retain non-English reviews while limiting the risk that translation alters the tone of customer statements. The domain-specific pretraining further makes the model well-suited for short, informal, and user-generated online reviews, which share many characteristics with online social media language. The use of a RoBERTa-based transformer model is further supported by Davoodi et al. (2026), who test several sentiment-classification models on Trustpilot e-commerce reviews and identify RoBERTa as the best-performing model.

The model predicts the probability that a text belongs to the negative, neutral, or positive sentiment class. It first produces raw class scores, which are transformed into probabilities. A continuous sentiment score is then constructed as the positive class probability minus the negative class probability, yielding values between -1 and 1. Higher values indicate more positive sentiment, and lower values indicate more negative sentiment. The neutral probability affects the score indirectly, as when a text is more neutral, both positive and negative probabilities tend to be lower, which moves the sentiment score closer to zero.

The customer sentiment measure is constructed from the full review text rather than from the review title. This is consistent with Cheung et al. (2008), who find that the most vital elements influencing usefulness and thereby consumer decision to adopt eWOM are information relevance and comprehensiveness. It also aligns with Chakraborty et al. (2022), who show that accurate sentiment classification benefits from models that capture both local phrase structure and broader contextual information, especially in longer sentences where polarity may shift across clauses. Taken together, these findings suggest that relying only on review titles may omit important sentiment information since titles provide only a condensed summary of the full review text. Sentiment extracted from the full review text is therefore used as the sentiment measure throughout the analysis.

The individual review sentiment score is aggregated at the firm-month level. We use review-volume weights as a proxy for brand importance, where the number of reviews reflects the scale of customer interaction and captures variation in customer engagement across brands. Accordingly, firm-month sentiment is computed as the mean of all review-level sentiment scores within a given month across the firm’s associated Trustpilot links. This approach implicitly assigns greater weights to brands with higher levels of customer activity and thereby approximates their relative importance within the parent firm’s operations. This is consistent with Jin (2023), where aggregation becomes review-weighted through the consolidation of product-level reviews. An alternative approach would be to use brand-level revenue weights as a proxy for each brand’s importance within the parent firm. However, consistent and comparable revenue data is not available across our sample, making this approach unfeasible. We therefore rely on review volume as the most practical proxy for brand importance.

In addition, the number of reviews underlying each monthly observation is recorded as a separate variable capturing review volume. This reflects the amount of consumer engagement and serves as an indicator of the reliability and informational content of the sentiment measure. The sentiment data is merged with the financial data, yielding a final dataset of 81 companies and 8,067 firm-month observations. Out of the firm-month observations, 5,864 observations contain review sentiment and form the basis of the empirical analysis. After controlling for public listing dates, the final dataset comprises 1,366,900 reviews.

3.3 Variable Definitions and Construction

The empirical models employ two dependent variables corresponding to the two types of firm outcomes examined, stock returns and firm fundamentals. In the return analysis, the dependent variable is next-month excess return. Excess return is defined as the firm’s monthly dividend-adjusted stock return minus the one-month risk-free rate. Consistent with Huang (2018), the firm-month sentiment measure is matched to stock returns in the subsequent month. This one-month lag ensures that sentiment is measured before the return outcome, reducing look-ahead bias. For the revenue growth analysis, the dependent variable is next-quarter log revenue growth. Quarterly sentiment is matched to revenue growth in the subsequent quarter, ensuring that customer sentiment is measured before

the operating outcome. This measure captures proportional changes in revenue rather than absolute changes, which are not directly comparable across firms of different sizes. It can therefore be interpreted approximately as percentage revenue growth.

The main explanatory variable used in the analysis is abnormal review sentiment, defined as the deviation of average firm-level sentiment in month t relative to its trailing twelve-month average, which serves as a firm-specific benchmark. The measure captures changes in consumer perceptions relative to recent historical sentiment, reflecting new information about the firm. The construction follows the abnormal review measure used by Jin (2023) and Huang (2018), who similarly define abnormal customer evaluations relative to historical firm-level benchmarks. To ensure that the sentiment measure is based on sufficient review activity, we impose a minimum threshold of five-reviews in the current month for which abnormal review sentiment is computed. The historical benchmark is based on the previous twelve-months and requires at least six-months within the twelve-month window to meet the same five review threshold. The rolling benchmark is then computed using all available valid sentiment observations within the twelve-month window. This approach balances data quality with sample retention by requiring sufficient historical review activity while allowing for occasional gaps in review coverage. Consequently, abnormal review sentiment for firm i in month t ($ARS_{i,t}$) is defined as:

$$ARS_{i,t} = S_{i,t} - \frac{1}{|H_{i,t}|} \sum_{k \in H_{i,t}} S_{i,t-k}$$

where $S_{i,t}$ denotes average review sentiment for firm i in month t , $H_{i,t} \subseteq \{1, \dots, 12\}$ denotes the set of valid lags in the previous twelve-month window, and $|H_{i,t}|$ denotes the number of valid lags.

Text-based sentiment serves as the primary review measure in the analysis, while abnormal review rating is employed as an alternative review-based variable in the robustness analysis. Following Huang (2018), and in line with the construction of abnormal review sentiment, abnormal review rating is defined as the firm-level average rating in month t minus the firm's average star rating over the previous twelve-month window. The same review threshold used for abnormal review sentiment is applied to abnormal review rating. Abnormal review rating for firm i in month t ($ARR_{i,t}$) is defined as:

$$ARR_{i,t} = R_{i,t} - \frac{1}{|H_{i,t}|} \sum_{k \in H_{i,t}} R_{i,t-k}$$

where $R_{i,t}$ denotes the average star rating for firm i in month t . The historical benchmark is constructed using the same set of valid lags, $H_{i,t}$, as in the abnormal review sentiment measure.

In addition to the main explanatory variables, the empirical models include both review-based and firm-level controls. The review-based controls, abnormal review count and dispersion, capture customer review activity and disagreement among customers, consistent with Jin (2023). We define abnormal review count as the deviation of the current month’s review count from its trailing twelve-month average. Unlike the abnormal review sentiment and rating measurement, no threshold is imposed on minimum review counts, as months with zero reviews provide valuable information about variation in customer activity. Abnormal review count for firm i in month t ($ARC_{i,t}$) is defined as:

$$ARC_{i,t} = N_{i,t} - \frac{1}{12} \sum_{k=1}^{12} N_{i,t-k}$$

where $N_{i,t}$ denotes the number of reviews for firm i in month t . The benchmark equals the average monthly review count over the previous twelve-months.

Sentiment dispersion is computed as the standard deviation of review-level sentiment scores within each firm-month, capturing dispersed opinions among customers. In consistency with sentiment dispersion, rating dispersion is defined as the standard deviation of star ratings within each firm-month.

For the revenue growth analysis, the monthly review variables are aggregated to the firm-quarter level as revenue is reported on a quarterly basis. Quarterly abnormal sentiment, quarterly abnormal review count, and quarterly review dispersion are calculated as the average of the corresponding monthly variables within each firm-quarter.

The firm-level control variables are selected based on prior studies of consumer reviews and financial outcomes. Huang (2018) controls for firm characteristics and established return predictors when examining whether abnormal customer ratings predict future stock returns, while Jin (2023) also accounts for firm characteristics when studying review tone and firm performance. Guided by this literature, we include firm size, book-to-market, momentum, leverage, and gross profitability as firm-level controls. These variables account for differences in expected returns, financial structure, and operating performance,

reducing the risk that the estimated relationship between abnormal review sentiment and the dependent variables reflects standard firm characteristics rather than customer-review sentiment.

Momentum is defined as the cumulative stock return from month $t-12$ to month $t-2$, thereby excluding the most recent month to avoid short-term return reversal effects. Book-to-market is defined as lagged common equity divided by market capitalization. Firm size is measured through market capitalization. Leverage is defined as lagged total debt divided by lagged total assets. Gross profitability is defined as lagged gross profit divided by lagged total assets.

3.4 Descriptive Statistics

Table 1
Summary Statistics - Monthly Panel

This table presents the summary statistics for the main variables used in the empirical analysis based on the monthly panel data, including review activity, review sentiment, star rating, return variables, and firm characteristics. The statistics are based on firm-month observations for Nordic listed retail firms between January 2014 to December 2025. We report the number of non-missing observations, mean, standard deviation, median, 25th percentile, 75th percentile, minimum, and maximum values for each variable. Market capitalization is reported in EURm.

Variable	N	Mean	Std	Median	P25	P75	Min	Max
Number of reviews	8,067	169.443	633.275	6.000	0.000	43.000	0.000	9,096.000
Average review sentiment	5,864	0.006	0.445	0.075	-0.332	0.351	-0.952	0.937
Abnormal review count	7,112	8.164	304.615	0.000	-2.250	2.833	-3,698.083	7,296.750
Abnormal review sentiment	3,498	0.008	0.201	0.010	-0.084	0.097	-0.943	1.065
Average star rating	5,864	3.117	1.330	3.308	1.989	4.306	1.000	5.000
Abnormal star rating	3,498	0.024	0.594	0.024	-0.221	0.280	-3.169	3.087
Review dispersion	5,219	0.553	0.186	0.577	0.465	0.663	0.000	1.326
Next-month excess return	7,986	0.007	0.159	-0.001	-0.057	0.060	-0.704	7.579
Momentum	7,033	0.117	0.723	0.027	-0.201	0.292	-0.965	25.518
Book-to-market	7,866	0.564	0.678	0.370	0.216	0.729	-2.581	11.018
Market capitalization	8,067	2,305.026	5,825.058	413.492	106.250	1,161.623	0.000	64,336.377
Leverage	7,703	0.274	0.176	0.265	0.149	0.373	0.000	0.999
Gross profitability	7,703	0.119	0.083	0.102	0.063	0.155	-0.120	0.673

Table 1 reports summary statistics for the main firm-month variables used in the monthly return analysis. The table presents the underlying review and return data prior to regression-specific transformation and winsorization. Quarterly variables used in the revenue growth analysis are reported separately under Section 4.3 in Table 9.

The number of reviews per firm-month exhibits substantial dispersion, with the mean number of reviews equaling 169.443, while the median equals only 6 and the 75th percentile

is 43. This indicates a highly skewed distribution in which most firm-months receive relatively few reviews, while a small subset attracts very high review volumes. The maximum of 9,096 reviews further shows that customer engagement is concentrated in a limited number of firm-month observations. This pattern is consistent with user-generated content platforms such as Trustpilot, where customer engagement tends to be concentrated among a limited number of firms or time periods. Abnormal review count also varies substantially, with a mean of 8.164 and a standard deviation of 304.615. This shows that firms occasionally experience large deviations in review activity relative to their own historical review volume, motivating the inclusion of abnormal review count as a review-based control variable.

Abnormal review sentiment has a mean of 0.008 and a median of 0.010, confirming that the variable is centered close to zero by construction. At the same time, the standard deviation of 0.201 indicates meaningful variation around firms' historical sentiment benchmarks. Therefore, the abnormal sentiment measure captures changes in customer evaluations relative to each firm's own recent sentiment level, rather than persistent differences in average sentiment across firms. Next-month excess returns have a mean of 0.007 and a median of -0.001. However, returns vary considerably across firm-month observations, with a standard deviation of 0.159 and values ranging from -0.704 to 7.579. The wide range reflects the presence of extreme return observations in the raw data.

The variation in the number of observations across variables reflects the construction of the main review measures. Average review sentiment and star ratings are available for 5,864 firm-month observations that satisfy the current-month review requirement. Abnormal review sentiment is available for 3,498 observations, as its construction further requires sufficient historical review coverage to compute the trailing twelve-month benchmark. Therefore, the decline in observations reflects the additional data requirements imposed by the construction of abnormal review sentiment, rather than ordinary missing data.

Table 2
Correlation Matrix - Monthly Panel

This table presents the pairwise correlations between the main variables used in the monthly empirical analysis. The statistics are based on monthly firm-level data for Nordic listed retail firms between January 2014 to December 2025.

	Abnormal review count	Abnormal sentiment	Abnormal rating	Dispersion	Next-month excess return	Momentum	B/M	MCAP	Leverage	GP
Abnormal review count	1.000									
Abnormal review sentiment	0.125	1.000								
Abnormal review rating	0.159	0.877	1.000							
Dispersion	0.002	0.094	0.071	1.000						
Next-month excess return	0.008	0.029	0.031	0.002	1.000					
Momentum	0.039	0.037	0.036	0.021	0.056	1.000				
Book-to-market	-0.044	-0.017	-0.012	-0.016	0.021	-0.216	1.000			
Market capitalization	0.023	0.014	0.018	-0.009	-0.016	0.009	-0.160	1.000		
Leverage	0.004	0.010	0.009	0.091	-0.019	-0.051	0.030	-0.091	1.000	
Gross profitability	0.027	0.015	0.027	0.050	0.022	0.043	-0.234	0.075	-0.255	1.000

Table 2 reports correlations for the main firm-month variables used in the monthly return analysis. The correlation between abnormal review sentiment and abnormal review rating is high, with a correlation of 0.877. This indicates that the text-based sentiment measure and star ratings capture closely related dimensions of customer evaluations. However, the measures are not identical, supporting the use of text-based sentiment as the primary measure and abnormal review rating as a robustness test. The correlations between the review-based variables and next-month excess returns are modest. Therefore, the correlation matrix does not indicate a strong pairwise association between customer review measures and subsequent returns. The remaining correlations are modest in magnitude, suggesting that multicollinearity is unlikely to be a general concern.

The descriptives reported above are based on the raw variables. Before estimating regressions, we make several additional adjustments to the variables used in the empirical analysis. First, we transform market capitalization by taking the natural logarithm of MCAP to reduce the influence of skewness in firm size. Observations with non-positive market capitalization are set to missing for the variable. Second, we set the book-to-market ratio to be missing when common equity or market capitalization is non-positive. Third, leverage and gross profitability are defined only when lagged total assets are positive. Finally, we winsorize abnormal review count, next-month excess returns, momentum, and book-to-market at the 0.1st and 99.9th percentiles, limiting the influence of extreme observations while preserving the sample. Since the set of required variables differs across model specifications, the effective sample varies across the empirical tests. We therefore allow the sample to vary by specification rather than imposing a common sample across all models. Each regression table reports the corresponding effective sample.

4. Empirical Analysis

Throughout the analysis, we employ two complementary econometric frameworks, Fama and MacBeth (1973) regressions and firm fixed-effects panel regressions. The Fama-MacBeth approach estimates period-by-period cross-sectional regressions and averages the estimated coefficients over time. This method tests whether firms with higher abnormal customer sentiment earn higher subsequent excess returns than other firms in the same month, thereby capturing cross-sectional return predictability. Inference is based on Newey and West (1987) standard errors, which are heteroskedasticity- and autocorrelation-consistent. This adjustment accounts for potential serial correlation in the monthly coefficient estimates. Fixed-effects regressions complement the Fama-MacBeth analysis by exploiting within-firm variation over time. These regressions test whether changes in a firm’s own abnormal customer sentiment are associated with subsequent changes in the dependent variable, either next-month excess returns or next-quarter revenue growth. Fixed-effects panel regressions absorb time-invariant differences across firms, while time fixed-effects absorb shocks common to all firms in the same period. Standard errors are clustered at the firm-level to allow for heteroskedasticity and serial correlation within firms over time.

The empirical analysis is structured around the three main research questions. First, we examine whether abnormal review sentiment predicts subsequent excess returns. This is tested using both Fama-MacBeth and fixed-effects panel regressions, and supplemented by a review-coverage split to assess whether the informativeness of sentiment is impacted by the number of reviews. Second, we investigate whether negative abnormal sentiment has stronger return relevance than positive abnormal sentiment. This is examined using both regression frameworks, along with additional tests incorporating review dispersion as a moderating factor. Third, we assess whether abnormal review sentiment predicts next-quarter revenue growth using fixed-effects panel regressions.

Together, these tests examine whether customer sentiment contains information relevant to stock returns, whether its return relevance is asymmetric between positive and negative sentiment, and how customer sentiment relates to subsequent operating performance.

4.1 Customer Sentiment and Next-Month Excess Return Predictability

We begin by examining whether customer sentiment predicts next-month excess returns. First, we examine whether abnormal review sentiment explains differences in next-month excess returns across firms by estimating the following Fama-MacBeth regression:

$$R_{i,t+1}^e = \alpha_t + \beta_t \cdot \text{ARS}_{i,t} + \boldsymbol{\gamma}'_t \mathbf{X}_{i,t} + \varepsilon_{i,t+1} \quad (1)$$

where $R_{i,t+1}^e$ denotes the excess return of firm i in month $t + 1$, $\text{ARS}_{i,t}$ is the firm's abnormal review sentiment in month t as defined in Section 3.3, and $\mathbf{X}_{i,t}$ is a vector of controls including abnormal review count, review dispersion, log market capitalization, book-to-market, momentum, leverage, and gross profitability.

Secondly, we examine whether changes in abnormal review sentiment within firms predict subsequent excess returns by estimating the following firm fixed-effects panel regression:

$$R_{i,t+1}^e = \alpha_i + \delta_t + \beta \cdot \text{ARS}_{i,t} + \boldsymbol{\gamma}' \mathbf{X}_{i,t} + \varepsilon_{i,t+1} \quad (2)$$

where α_i and δ_t denote firm and month fixed-effects, respectively, and all other variables are as defined in Equation (1).

Lastly, the sample is split by review coverage to assess whether the effect of sentiment on next-month excess returns depends on the number of reviews.

Table 3
Fama-MacBeth Regressions of Excess Returns

This table presents Fama-MacBeth cross-sectional regressions of next-month excess returns on abnormal review sentiment and control variables. Column (i) presents the baseline specification without additional controls. Column (ii) includes review-related controls, specifically abnormal review count and dispersion. Column (iii) adds standard firm-level controls, including market capitalization, book-to-market ratio and momentum. Column (iv) further extends the specification by including leverage and gross profitability. We require a minimum of 15 firms in a month for the cross-sectional regression to be included. Newey-West standard errors are reported in parentheses.

	(i) No controls	(ii) Review controls	(iii) Standard controls	(iv) Extended controls
Abnormal review sentiment	0.0173 (0.0150)	0.0195 (0.0151)	0.0213 (0.0154)	0.0172 (0.0154)
Abnormal review count		<0.0001 ² (<0.0001) ⁵	<0.0001 ³ (<0.0001) ⁶	<0.0001 ⁴ (<0.0001) ⁷
Review dispersion		0.0016 (0.0205)	0.0090 (0.0219)	0.0082 (0.0244)
Log MCAP			-0.0023 (0.0015)	-0.0020 (0.0018)
Book-to-market			-0.0010 (0.0081)	0.0034 (0.0085)
Momentum			0.0312*** (0.0096)	0.0300*** (0.0094)
Leverage				-0.0156 (0.0149)
Gross profitability				0.0441 (0.0474)
Months	77	77	77	77
Avg. firms per month	39.81	39.81	38.66	38.23
Avg. R^2	0.034	0.109	0.254	0.313

, **, and * indicate statistical significance at the 10%, 5%, and 1% levels, respectively.*

Table 3 reports that the coefficients on abnormal review sentiment are positive across all Fama-MacBeth specifications, ranging from 0.0172 to 0.0213. The directional pattern indicates that firms with higher abnormal customer sentiment earn higher subsequent excess returns. However, none of these coefficients are statistically significant, suggesting that the positive association is not estimated with sufficient precision to support a reliable return-predictive effect across firms. Therefore, the results provide limited evidence that abnormal customer sentiment meaningfully explains cross-sectional variation in next-month excess returns. The inclusion of control variables increases average R^2 across

² 1.45×10^{-6} ³ 7.31×10^{-7} ⁴ 6.89×10^{-6} ⁵ 7.23×10^{-6} ⁶ 7.81×10^{-6} ⁷ 9.03×10^{-6}

specifications, indicating improved explanatory power. In the specification that includes momentum, the momentum coefficient is positive and statistically significant at the 1% level, consistent with its role as an established return predictor.

Table 4
Fixed-Effects Panel Regressions of Excess Returns

This table presents the results from fixed-effects panel regressions of next-month excess returns on abnormal review sentiment and control variables with firm and month fixed-effects. Column (i) presents the baseline specification without additional controls. Column (ii) includes review-related controls, specifically abnormal review count and dispersion. Column (iii) adds standard firm-level controls, including market capitalization, book-to-market ratio and momentum. Column (iv) further extends the specification by including leverage and gross profitability. Firm-clustered standard errors are reported in parentheses.

	(i) No controls	(ii) Review controls	(iii) Standard controls	(iv) Extended controls
Abnormal review sentiment	0.0080 (0.0115)	0.0097 (0.0113)	0.0067 (0.0114)	0.0055 (0.0114)
Abnormal review count		$>-0.0001^8$ (<0.0001) ¹¹	$>-0.0001^9$ (<0.0001) ¹²	$>-0.0001^{10}$ (<0.0001) ¹³
Review dispersion		-0.0086 (0.0213)	-0.0047 (0.0206)	-0.0052 (0.0210)
Log MCAP			-0.0468*** (0.0086)	-0.0506*** (0.0092)
Book-to-market			-0.0027 (0.0090)	-0.0057 (0.0084)
Momentum			0.0144** (0.0060)	0.0144** (0.0061)
Leverage				-0.0622* (0.0363)
Gross profitability				-0.0389 (0.0304)
Months	131	131	130	130
Avg. firms per month	68	68	67	67
Adjusted R ²	0.198	0.198	0.229	0.231

, **, and * indicate statistical significance at the 10%, 5%, and 1% levels, respectively.*

Table 4 reports the fixed-effects panel regressions of next-month excess returns, in which the coefficients on abnormal review sentiment are positive ranging from 0.0055 to 0.0097, but remain statistically insignificant across all specifications. Therefore, the results provide limited evidence that within-firm changes in abnormal customer sentiment predict higher subsequent excess returns following periods of abnormally positive sentiment. The

⁸ -5.03×10^{-6} ⁹ -3.97×10^{-6} ¹⁰ -3.41×10^{-6} ¹¹ 5.17×10^{-6} ¹² 6×10^{-6} ¹³ 5.89×10^{-6}

momentum variable remains positive and statistically significant at the 5% level, indicating that the model captures an established predictor of returns. The adjusted R^2 increases from 0.198 in the no control specification (i) to 0.231 in the extended control specification (iv), indicating that the additional controls improve explanatory power.

As an extension to the main analysis, we finally examine whether the relation between abnormal review sentiment and subsequent excess returns varies with review coverage. Review coverage reflects the amount of underlying customer information used to construct the sentiment measure. Sentiment based on a larger number of reviews may provide a more representative signal of customer perceptions, while sentiment based on fewer reviews may be more affected by noise. This is consistent with the view that aggregating opinions across a broader set of individuals can improve signal reliability (Surowiecki, 2004; Chen et al., 2014). The skewed distribution of review activity in the sample further motivates this analysis. To examine whether the relation between sentiment and next-month excess returns differs across coverage levels, firms are classified as high-coverage if their number of reviews exceeds the median across all firms in a given month, and as low-coverage otherwise.

Table 5
Fama-MacBeth Regressions of Excess Returns by Review Coverage

This table presents the results from Fama-MacBeth cross-sectional regressions of next-month excess returns on abnormal review sentiment, separately for low- and high-coverage observations. Columns (i) - (iv) report low-coverage observations, while columns (v) - (viii) report high-coverage observations. Within each group, the specifications progressively add controls: no controls, review-related controls, standard firm-level controls, and extended firm-level controls. Review-related controls include abnormal review count and review dispersion. Standard firm-level controls include market capitalization, book-to-market ratio, and momentum. Extended controls further include leverage and gross profitability. We require a minimum of 15 firms in a month for the cross-sectional regression to be included. Newey-West standard errors are reported in parentheses.

	Low-Coverage				High-Coverage			
	(i) No controls	(ii) Review controls	(iii) Standard controls	(iv) Extended controls	(v) No controls	(vi) Review controls	(vii) Standard controls	(viii) Extended controls
Abnormal review sentiment	0.0041 (0.0170)	0.0024 (0.0184)	0.0063 (0.0235)	-0.0113 (0.0351)	0.0292 (0.0180)	0.0342* (0.0179)	0.0333 (0.0208)	0.0338 (0.0216)
Abnormal review count		-0.0010* (0.0006)	-0.0009 (0.0008)	-0.0010 (0.0009)		>-0.0001 ¹⁴ (<0.0001) ¹⁷	<0.0001 ¹⁵ (<0.0001) ¹⁸	<0.0001 ¹⁶ (<0.0001) ¹⁹
Review dispersion		0.0188 (0.0281)	0.0213 (0.0272)	0.0146 (0.0345)		-0.0186 (0.0242)	-0.0034 (0.0270)	-0.0061 (0.0317)
Log MCAP			-0.0022 (0.0034)	-0.0037 (0.0048)			-0.0026 (0.0017)	-0.0021 (0.0020)
Book-to-market			0.0145 (0.0124)	0.0127 (0.0137)			-0.0031 (0.0089)	0.0037 (0.0098)
Momentum			0.0275** (0.0111)	0.0285** (0.0132)			0.0346*** (0.0117)	0.0317*** (0.0108)
Leverage				0.0077 (0.0275)				-0.0197 (0.0195)
Gross profitability				-0.0089 (0.0798)				0.0225 (0.0547)
Months	29	29	29	29	77	77	77	77
Avg. firms per month	18.90	18.90	18.76	18.28	29.22	29.22	28.17	27.92
Avg. R ²	0.065	0.179	0.388	0.518	0.044	0.132	0.310	0.384

, **, and * indicate statistical significance at the 10%, 5%, and 1% levels, respectively.*

Table 5 shows that the Fama-MacBeth results reveal a difference between high- and low-coverage observations. In the high-coverage sample, the coefficient on abnormal review sentiment is positive across all four specifications, ranging from 0.0292 to 0.0342. The coefficient reaches marginal significance at the 10% level in the review controls specification (vi), while the remaining specifications fall slightly short of the 10% significance threshold²⁰. This provides suggestive evidence that abnormal review sentiment may be positively associated with next-month excess returns across firms when the sentiment measure

¹⁴ -1.59×10^{-7} ¹⁵ 1.32×10^{-6} ¹⁶ 7.97×10^{-6} ¹⁷ 7.75×10^{-6} ¹⁸ 8.37×10^{-6} ¹⁹ 9.8×10^{-6} ²⁰ With Newey-West adjusted p-values of 10.43%, 5.53%, 10.89% and 11.80% respectively for each specification.

is based on a larger number of reviews. In contrast, the low-coverage sample shows substantially weaker results. The coefficients on abnormal review sentiment are close to zero, ranging from -0.0113 to 0.0063, and remain statistically insignificant across all specifications. Therefore, the review coverage analysis provides suggestive but not robust evidence that customer sentiment is more informative when based on a larger number of reviews across firms. Among the control variables, momentum is positive and statistically significant in both coverage samples, consistent with its role as an established return predictor. The average R^2 increases with the inclusion of additional controls, indicating that the controls improve explanatory power.

The corresponding fixed-effects panel regressions, presented in Table A.1 in the Appendix, show a similar directional pattern but weaker evidence than the Fama-MacBeth results. The coefficients on abnormal review sentiment are positive in both the high- and low-coverage samples, but remain statistically insignificant across all specifications. Similar to the cross-sectional findings, the estimates are generally larger in the high-coverage sample, suggesting that sentiment based on more reviews may provide a stronger signal. However, because none of the fixed-effects estimates reach statistical significance, the coverage does not provide reliable within-firm evidence.

Overall, the findings provide limited support for H1, which predicts that firms with higher abnormal customer review sentiment experience higher subsequent excess returns. The coefficient on abnormal review sentiment is positive across both the Fama-MacBeth and fixed-effects specifications, consistent with the predicted direction. Despite the directional consistency, the coefficients are not statistically significant in either the cross-sectional or within-firm setting. The review coverage analysis adds nuance to this finding. The Fama-MacBeth results suggest that sentiment may be more informative when review coverage is high, as reflected in larger coefficients and marginal significance in the high-coverage sample. However, this pattern is not robust across specifications and does not hold in the fixed-effects model, where sentiment remains insignificant regardless of coverage level.

Taken together, these findings indicate that the predictive power of customer sentiment is weak and not robust across empirical specifications. While there is some evidence that sentiment may contain information in settings with high review coverage, this effect is not consistent across methods and does not provide sufficient support for the conclusion that

customer sentiment reliably predicts future stock returns.

4.2 Asymmetric Return Effects of Positive and Negative Customer Sentiment

We proceed with our analysis by examining the asymmetric effects of abnormal customer sentiment on next-month excess returns. Consistent with Jin (2023), we decompose abnormal review sentiment into two components for each month. The positive component captures sentiment above the firm’s prior benchmark and equals abnormal review sentiment when it is positive, and zero otherwise. The negative component captures sentiment below the firm’s prior benchmark and equals the absolute value of abnormal review sentiment when it is negative, and zero otherwise. Therefore, a negative coefficient on this variable implies that more negative abnormal review sentiment is associated with lower subsequent excess returns.

Table 6
Summary Statistics for Asymmetric Sentiment Variables

This table presents summary statistics for the asymmetric decomposition of abnormal review sentiment into positive and negative components. Share non-zero reports the fraction of observations where the variable is different from zero.

Variable	N	Share non-zero (%)	Mean	Std	Median	P95	Max
Positive abnormal review sentiment	1,873	53.5	0.138	0.146	0.088	0.444	1.065
Negative abnormal review sentiment	1,625	46.5	0.143	0.143	0.096	0.426	0.943

Table 6 presents summary statistics for the asymmetry decomposition. Positive abnormal review sentiment accounts for 1,873 observations, corresponding to 53.5% of the sample, while negative abnormal review sentiment accounts for 1,625 observations, corresponding to 46.5% of the sample. Therefore, the two components are relatively balanced in frequency. The two components are also similar in magnitude, suggesting that the asymmetry tests are not driven by a large imbalance in the number or size of positive and negative sentiment deviations.

To assess whether the effect of abnormal review sentiment on subsequent excess returns is asymmetric, we conduct a series of tests. First, we examine whether positive and negative abnormal review sentiment exhibit different associations with subsequent excess returns across firms by estimating the following Fama-MacBeth regression:

$$R_{i,t+1}^e = \alpha_t + \beta_t^+ \cdot ARS_{i,t}^+ + \beta_t^- \cdot ARS_{i,t}^- + \gamma_t' \mathbf{X}_{i,t} + \varepsilon_{i,t+1} \quad (3)$$

where $ARS_{i,t}^+ = \max(ARS_{i,t}, 0)$ captures positive abnormal sentiment, and $ARS_{i,t}^- = \max(-ARS_{i,t}, 0)$ captures the magnitude of negative abnormal sentiment. The vector $\mathbf{X}_{i,t}$ contains the control variables defined in Equation (1).

Secondly, we examine whether changes in positive and negative abnormal review sentiment within firms predict subsequent excess returns over time by estimating the following firm fixed-effects panel regression:

$$R_{i,t+1}^e = \alpha_i + \delta_t + \beta^+ \cdot ARS_{i,t}^+ + \beta^- \cdot ARS_{i,t}^- + \gamma' \mathbf{X}_{i,t} + \varepsilon_{i,t+1} \quad (4)$$

where α_i and δ_t denote firm and month fixed-effects, respectively, and the positive and negative sentiment components are defined as in Equation (3). The vector $\mathbf{X}_{i,t}$ contains the control variables defined in Equation (1).

Third, we test whether review dispersion moderates the return relevance of negative sentiment. The Fama-MacBeth specification is presented in Equation (5) while the corresponding firm fixed-effects specification is presented in Equation (6):

$$\begin{aligned} R_{i,t+1}^e = & \alpha_t + \beta_t^+ \cdot ARS_{i,t}^+ + \beta_t^- \cdot ARS_{i,t}^- + \theta_t \cdot \text{DispersionStd}_{i,t} \\ & + \lambda_t \cdot (ARS_{i,t}^- \times \text{DispersionStd}_{i,t}) + \gamma_t' \mathbf{X}_{i,t} + \varepsilon_{i,t+1} \end{aligned} \quad (5)$$

$$\begin{aligned} R_{i,t+1}^e = & \alpha_i + \delta_t + \beta^+ \cdot ARS_{i,t}^+ + \beta^- \cdot ARS_{i,t}^- + \theta \cdot \text{DispersionStd}_{i,t} \\ & + \lambda \cdot (ARS_{i,t}^- \times \text{DispersionStd}_{i,t}) + \gamma' \mathbf{X}_{i,t} + \varepsilon_{i,t+1} \end{aligned} \quad (6)$$

where Equation (5) is estimated using the Fama-MacBeth approach and Equation (6) using a firm fixed-effects panel specification with firm fixed-effects α_i and month fixed-effects δ_t . In both specifications, $\text{DispersionStd}_{i,t}$ denotes review dispersion standardized cross-sectionally each month, and λ captures the extent to which review dispersion moderates the return relevance of negative sentiment. The positive and negative sentiment components are defined as in Equation (3) and the control vector $\mathbf{X}_{i,t}$ is defined as in Equation (1).

Table 7
Fama-MacBeth Regressions on Asymmetric Sentiment and Excess Returns

This table presents Fama-MacBeth cross-sectional regressions of next-month excess returns on positive and negative abnormal review sentiment and control variables. Column (i) presents the baseline specification without additional controls. Column (ii) includes review-related controls, specifically abnormal review count and dispersion. Column (iii) adds standard firm-level controls, including market capitalization, book-to-market ratio, and momentum. Column (iv) further extends the specification by including leverage and gross profitability. $\beta_{\text{pos}} - \beta_{\text{neg}}$ reports the difference between the positive and negative sentiment coefficients. We require a minimum of 15 firms in a month for the cross-sectional regression to be included. Newey-West standard errors are reported in parentheses.

	(i) No controls	(ii) Review controls	(iii) Standard controls	(iv) Extended controls
Positive abnormal sentiment	0.0197 (0.0224)	0.0268 (0.0225)	0.0113 (0.0264)	0.0153 (0.0252)
Negative abnormal sentiment	-0.0408* (0.0229)	-0.0517 (0.0356)	-0.0954* (0.0528)	-0.0802 (0.0518)
$\beta_{\text{pos}} - \beta_{\text{neg}}$	0.0605* (0.0343)	0.0785* (0.0471)	0.1067 (0.0650)	0.0955 (0.0633)
Abnormal review count		<0.0001 ²¹ (<0.0001) ²⁴	<0.0001 ²² (<0.0001) ²⁵	<0.0001 ²³ (<0.0001) ²⁶
Review dispersion		0.0029 (0.0210)	0.0124 (0.0225)	0.0128 (0.0242)
Log MCAP			-0.0031 (0.0020)	-0.0025 (0.0022)
Book-to-market			-0.0040 (0.0092)	0.0015 (0.0094)
Momentum			0.0302*** (0.0108)	0.0293*** (0.0107)
Leverage				-0.0154 (0.0156)
Gross profitability				0.0365 (0.0432)
Months	77	77	77	77
Avg. firms per month	39.81	39.81	38.66	38.23
Avg. R ²	0.063	0.143	0.289	0.349

, **, and * indicate statistical significance at the 10%, 5%, and 1% levels, respectively.*

Table 7 shows that the coefficients of positive abnormal sentiment remain positive across all model specifications, ranging from 0.0113 to 0.0268. However, the estimates are statistically insignificant across all specifications, suggesting that abnormal positive customer sentiment does not reliably predict higher subsequent excess returns. In contrast,

²¹ 1.36×10^{-6} ²² 2.56×10^{-6} ²³ 7.85×10^{-6} ²⁴ 7.37×10^{-6} ²⁵ 7.67×10^{-6} ²⁶ 8.79×10^{-6}

the coefficient on negative abnormal sentiment is negative across all specifications, ranging from -0.0408 to -0.0954. The coefficient is marginally significant at the 10% level in the no controls (i) and standard controls (iii) specifications, suggesting that firms experiencing abnormally negative customer sentiment may earn lower subsequent excess returns. The average R^2 increases with the inclusion of additional controls, indicating improved explanatory power. In the specifications that include momentum, the momentum coefficient is positive and statistically significant at the 1% level, consistent with its role as an established return predictor.

The coefficient difference test ($\beta_{\text{pos}} - \beta_{\text{neg}}$) shows whether the coefficients on positive and negative sentiment differ significantly from each other. The difference between the coefficients is positive and marginally significant at the 10% level in the no control (i) and review control (ii) specifications, suggesting that negative abnormal sentiment has a stronger adverse return association than positive abnormal sentiment in these models. However, the difference becomes statistically insignificant after adding standard and extended controls. The cross-sectional results, therefore, provide some evidence that negative sentiment has a stronger return association than positive sentiment across firms, but the evidence is not robust across all specifications.

The fixed-effects results provide weaker evidence of asymmetry, as shown in Table A.2 in the Appendix. The coefficient on negative abnormal sentiment remains negative across specifications, but neither the positive nor the negative sentiment coefficients are statistically significant. The coefficient difference test also provides no evidence that negative sentiment has stronger return relevance than positive sentiment within firms over time. These results indicate that within-firm changes in customer sentiment do not translate into reliably different subsequent excess returns following abnormally positive or negative sentiment periods.

To provide additional insights into the main asymmetry result, we further examine whether review dispersion moderates the return relevance of negative customer sentiment. Review dispersion captures the degree of disagreement across customer reviews, where higher dispersion indicates greater inconsistency in customer experiences and increased uncertainty about the firm’s overall customer perception. Babić Rosario et al. (2016) find that greater eWOM variability is associated with unfavorable consumer outcomes, as higher disagreement among reviewers increases functional risk. To assess whether

this uncertainty matters in the return setting, we interact the negative sentiment with standardized review dispersion. A negative interaction coefficient indicates that the return penalty associated with negative abnormal sentiment becomes stronger as review dispersion increases.

Table 8
Fama-MacBeth Regressions on Negative Sentiment, Review Dispersion, and Subsequent Excess Returns

This table presents Fama-MacBeth cross-sectional regressions of next-month excess returns on positive abnormal sentiment, negative abnormal sentiment magnitude, review dispersion, and the interaction between negative sentiment and dispersion. Review dispersion is standardized within month before constructing the interaction term, so zero corresponds to the monthly mean. Accordingly, the coefficient on negative abnormal sentiment magnitude is evaluated at mean dispersion, while the coefficient on review dispersion is evaluated when negative abnormal sentiment magnitude equals zero. Column (i) presents the baseline specification without additional controls. Column (ii) includes review-related controls, specifically abnormal review count and dispersion. Column (iii) adds standard firm-level controls, including market capitalization, book-to-market ratio, and momentum. Column (iv) further extends the specification by including leverage and gross profitability. We require a minimum of 15 firms in a month for the cross-sectional regression to be included. Newey-West standard errors are reported in parentheses.

	(i) No controls	(ii) Review controls	(iii) Standard controls	(iv) Extended controls
Negative abnormal sentiment	-0.0433 (0.0302)	-0.0426 (0.0311)	-0.0991** (0.0491)	-0.0923* (0.0507)
Positive abnormal sentiment	0.0169 (0.0237)	0.0227 (0.0234)	0.0053 (0.0278)	0.0119 (0.0271)
Review dispersion	0.0072 (0.0043)	0.0062 (0.0042)	0.0097* (0.0049)	0.0101* (0.0053)
Negative sentiment $\times$ dispersion	-0.1134** (0.0547)	-0.1035* (0.0556)	-0.1422** (0.0695)	-0.1496** (0.0603)
Abnormal review count		$<0.0001^{27}$ (<0.0001) ³⁰	$<0.0001^{28}$ (<0.0001) ³¹	$<0.0001^{29}$ (<0.0001) ³²
Log MCAP			-0.0033 (0.0022)	-0.0026 (0.0024)
Book-to-market			-0.0014 (0.0094)	0.0055 (0.0098)
Momentum			0.0287*** (0.0106)	0.0308** (0.0121)
Leverage				-0.0235 (0.0171)
Gross profitability				0.0289 (0.0409)
Months	77	77	77	77
Avg. firms per month	39.81	39.81	38.66	38.23
Avg. R ²	0.132	0.173	0.321	0.376

, **, and * indicate statistical significance at the 10%, 5%, and 1% levels, respectively.*

²⁷ 6.79×10^{-7} ²⁸ 3.58×10^{-6} ²⁹ 7.24×10^{-6} ³⁰ 7.35×10^{-6} ³¹ 7.93×10^{-6} ³² 8.80×10^{-6}

Table 8 shows that the interaction between negative abnormal sentiment and review dispersion is negative across all specifications, with coefficients ranging from -0.1035 to -0.1496. The interaction term is statistically significant at the 5% level in the no control (i), standard control (iii), and extended control (iv) specifications, and marginally significant at the 10% level in the review controls (ii) specification. This indicates that the negative association between abnormal negative sentiment and subsequent excess returns is stronger when review dispersion is higher in the cross-sectional setting. In other words, negative sentiment appears more return-relevant when customer opinions are more dispersed. The stand alone coefficient on negative abnormal sentiment is negative across all specifications, but the interaction term provides the central test of whether dispersion moderates the return relevance of negative sentiment. The inclusion of control variables increases average R^2 across specifications, indicating improved explanatory power. In the specification that includes momentum, the momentum coefficient is positive and statistically significant, consistent with its role as an established return predictor.

The corresponding fixed-effects panel regressions, shown in Table A.3 in the Appendix, report negative interaction coefficients across all specifications, but none are statistically significant. This suggests that the dispersion moderation effect does not provide reliable within-firm evidence. Taken together, the results indicate that review dispersion amplifies the cross-sectional return relevance of negative abnormal review sentiment, while the fixed-effects results provide weaker evidence that the same relation holds within firms over time.

Overall, the results provide limited support for H2, which predicts that negative abnormal customer sentiment exhibits a stronger association with subsequent excess returns than positive abnormal sentiment. In the Fama-MacBeth regressions, positive abnormal sentiment remains statistically insignificant, while negative abnormal sentiment is negative across all specifications and marginally significant in selected models. This pattern suggests that negative sentiment may have a more consistent association with subsequent returns than positive sentiment across firms, although not robust across specifications. However, the coefficient difference tests provide only weak evidence of asymmetry, as the difference between the positive and negative sentiment coefficients is marginally significant in the simpler specifications but becomes statistically insignificant after adding standard and extended controls. The fixed-effects regressions provide no

evidence that this asymmetric pattern exists within firms over time.

The interaction test provides stronger evidence that the return relevance of negative sentiment depends on review dispersion. In the Fama-MacBeth specifications, the interaction between negative abnormal sentiment and review dispersion is negative and marginally or statistically significant across all specifications. This indicates that negative sentiment is more strongly associated with lower subsequent returns across firms when customer opinions are more dispersed. However, the corresponding fixed-effects estimates are not statistically significant, suggesting that this moderation effect is primarily cross-sectional rather than within-firm over time.

Taken together, the evidence suggests that negative abnormal sentiment is more closely related to subsequent returns than positive abnormal sentiment in cross-sectional settings characterized by high review dispersion. However, the results do not provide robust support for a general asymmetric return effect. The asymmetric effect is therefore only partially supported, and the evidence should be interpreted as suggestive rather than conclusive.

4.3 Customer Sentiment as a Predictor of Future Firm Fundamentals

We conclude our empirical analysis with a fixed-effects panel regression to estimate the impact of abnormal review sentiment on future firm fundamentals. As described in Section 3.3, the revenue growth analysis aggregates review variables to the firm-quarter level as revenue is reported at the quarterly basis. The dependent variable is next-quarter log revenue growth, capturing top-line performance, and provides a measure of whether sentiment translates into future sales. The main explanatory variable is quarterly abnormal review sentiment constructed from the monthly abnormal review sentiment measure.

Table 9
Summary Statistics for Main Quarterly Variables

This table presents summary statistics for the main quarterly variables used in the future revenue growth analysis. All variables are constructed at the firm-quarter level computed as the average of the months within each quarter. Next-quarter log revenue growth is defined as the log change in revenue from one quarter to the next. Next-quarter log revenue growth and quarterly abnormal review count are based on values winsorized at the 0.1st and 99.9th percentiles.

Variable	N	Mean	Std	P25	Median	P75
Next-quarter log revenue growth	1,138	0.0136	0.3034	-0.1126	0.0178	0.1300
Quarterly abnormal sentiment	1,138	0.0068	0.1585	-0.0625	0.0043	0.0744
Quarterly abnormal review count	1,138	8.8735	326.7793	-8.6736	0.6806	12.8819
Quarterly review dispersion	1,138	0.5727	0.1119	0.5009	0.5839	0.6523

Table 9 shows that next-quarter log revenue growth has a mean of 0.0136 and a median of 0.0178, indicating that the average firm-quarter in the regression sample experiences positive next-quarter revenue growth. However, the standard deviation of 0.3034 and the interquartile range from -0.1126 to 0.1300 show that future revenue growth varies substantially across firms. The mean of quarterly abnormal sentiment is close to zero, consistent with the construction of abnormal sentiment.

Whereas the previous analysis examined the predictive power of abnormal review sentiment for stock returns, this analysis examines whether abnormal review sentiment contains information about firms' operating performance. Specifically, we test predictions of a firm's future revenue growth. Unlike the return analysis, the revenue growth test is estimated only using a firm fixed-effects panel regression. This reflects that revenue growth is primarily an operating outcome shaped by firm-specific characteristics and baseline growth patterns. The fixed-effects model therefore tests whether abnormal sentiment predicts changes in future revenue growth within the same firm, while controlling for persistent differences across firms and common time shocks, using the following specification:

$$\Delta \text{LogRev}_{i,q+1} = \alpha_i + \delta_q + \beta \cdot \text{ARS}_{i,q} + \gamma' \mathbf{X}_{i,q} + \varepsilon_{i,q+1} \quad (7)$$

where $\Delta \text{LogRev}_{i,q+1}$ is the log change in revenue from quarter q to quarter $q + 1$, $\text{ARS}_{i,q}$ is quarterly abnormal review sentiment computed as the average of the monthly measure within firm-quarter q , α_i and δ_q are firm and quarter fixed-effects, respectively, and $\mathbf{X}_{i,q}$ contains the quarterly analogues of the controls in Equation (1).

Table 10
**Fixed-Effects Panel Regressions on Abnormal Customer
Sentiment and Future Revenue Growth**

This table presents the results from panel regressions of next-quarter log revenue growth on abnormal review sentiment and control variables with firm and quarter fixed-effects. Column (i) presents the baseline specification without additional controls. Column (ii) includes review-related controls, specifically abnormal review count and dispersion. Column (iii) adds standard firm-level controls, including market capitalization, book-to-market ratio, and momentum. Column (iv) further extends the specification by including leverage and gross profitability. Firm-clustered standard errors are reported in parentheses.

	(i) No controls	(ii) Review controls	(iii) Standard controls	(iv) Extended controls
Quarterly abnormal sentiment	0.0381 (0.0308)	0.0707** (0.0349)	0.0745** (0.0353)	0.0769** (0.0353)
Quarterly abnormal review count		-0.0001*** (<0.0001) ³³	-0.0001*** (<0.0001) ³⁴	-0.0001*** (<0.0001) ³⁵
Quarterly review dispersion		0.0398 (0.1071)	0.0216 (0.1090)	0.0095 (0.1084)
Log MCAP			-0.0247 (0.0194)	-0.0211 (0.0178)
Book-to-market			-0.0218 (0.0271)	-0.0244 (0.0276)
Momentum			0.0080 (0.0153)	0.0191 (0.0151)
Leverage				0.1321 (0.1726)
Gross profitability				-0.6805** (0.3448)
Quarters	43	43	43	43
Avg. firms per quarter	66	66	66	66
Adjusted R ²	0.093	0.104	0.101	0.113

, **, and * indicate statistical significance at the 10%, 5%, and 1% levels, respectively.*

Table 10 shows that quarterly abnormal sentiment is positively associated with future revenue growth across all model specifications. The coefficient on quarterly abnormal sentiment increases from 0.0381 in the no control specification (i) to 0.0769 in the extended control specification (iv). While the no control specification (i) is statistically insignificant, the coefficient becomes statistically significant at the 5% level once controls are included. These results suggest that firms with higher abnormal customer sentiment experience higher future revenue growth. The adjusted R^2 is generally higher in the controlled specifications

Among the control variables, the coefficient on quarterly abnormal review count is

³³ 2.77×10^{-5} ³⁴ 2.85×10^{-5} ³⁵ 3.55×10^{-5}

consistently negative and statistically significant at the 1% level, although economically small in magnitude, with a coefficient of -0.0001 across all specifications. This suggests that unusually high review volume is associated with lower future revenue growth after controlling for sentiment. Furthermore, gross profitability is negative and statistically significant at the 5% significance level, suggesting that firms with higher gross profitability exhibit lower subsequent revenue growth within the sample, conditional on the other controls.

Overall, the results support H3, which predicts that firms with higher abnormal customer sentiment experience higher future revenue growth. Quarterly abnormal customer sentiment positively predicts future revenue growth, and the relationship becomes statistically significant at the 5% level after the inclusion of control variables. This means that firms experiencing customer sentiment above their recent historical benchmark tend to exhibit higher revenue growth in the subsequent quarter. Although the results show that abnormal sentiment predicts future revenue growth, they do not establish causality that changes in customer sentiment cause changes in revenue.

5. Robustness

To assess the robustness of the main results, we conduct two sensitivity analyses. First, we replace text-based abnormal review sentiment with abnormal star ratings. Although the two measures are highly correlated, they may capture different dimensions of customer evaluations. This test therefore examines whether the main findings are robust to using an alternative review-based measure. Second, we test for threshold dependence by increasing the minimum review requirement from five to ten reviews when constructing abnormal review sentiment. This doubles the review threshold, imposing a substantially stricter data requirement. While the higher threshold may reduce measurement noise by requiring more review information per firm-month, it also reduces sample size. The test therefore examines whether the results persist in a more restrictive, but potentially cleaner, sample.

The star rating robustness test produces results that are directionally consistent with the main analysis, but generally weaker, as shown in Table A.4 in the Appendix. The return predictability coefficients remain positive in the Fama-MacBeth specifications, but they are statistically insignificant. The asymmetry tests also show a similar directional

pattern, with negative abnormal ratings generally more associated with subsequent returns than positive abnormal ratings. However, the statistical evidence remains weak and does not provide strong support for asymmetry. The interaction between negative abnormal ratings and review dispersion is statistically insignificant across specifications, unlike the stronger interaction result in the main text-sentiment analysis. The revenue growth test also weakens when abnormal star ratings replace text-based sentiment. Although the coefficients remain positive, they are no longer statistically significant. Taken together, these results suggest that the main findings are sensitive to the choice of review measure. One possible explanation is that numerical star ratings may not fully capture the multiple dimensions of customer experiences, whereas text-based sentiment can reflect richer and more nuanced customer evaluations. This is consistent with prior evidence that textual content captures dimensions of customer evaluations that numerical ratings alone do not fully capture (Archak et al., 2011; Jin, 2023).

The robustness tests for the 10-review threshold further support the directional findings of the main results, as highlighted in Table A.5 in the Appendix. The return predictability coefficients remain positive in the Fama-MacBeth specifications, and the review-control specification reaches marginal statistical significance. However, the evidence from the high-coverage split weakens relative to the main analysis. The asymmetry tests also become weaker under the stricter threshold. Negative abnormal sentiment generally retains the expected negative sign in most specifications, but the coefficients are not statistically significant, and the coefficient difference tests provide only limited evidence of asymmetry. Furthermore, the interaction between negative abnormal sentiment and review dispersion remains negative across specifications, but becomes statistically insignificant. One possible explanation is that the stricter review threshold reduces the number of eligible firm-month observations, lowering statistical precision despite the similar directional patterns. In contrast, the revenue growth results remain robust. Abnormal customer sentiment remains positively and significantly associated with future revenue growth after the inclusion of controls. This suggests that the firm-fundamentals result is less sensitive to the review threshold than the return-based results.

Taken together, the robustness tests do not overturn the main findings. The coefficients generally retain the same direction as in the main analysis, although the statistical strength varies. Overall, the evidence is strongest and most consistent for future revenue growth,

while the return-based findings should be interpreted with greater caution.

6. Discussion

The first research question sets out to answer whether abnormal sentiment from company-level customer reviews predicts subsequent excess returns. The results provide limited support for this relation. Across both the Fama-MacBeth and fixed-effects specifications, the coefficients on abnormal review sentiment are consistently directionally positive but do not reach statistical significance. This indicates that higher abnormal customer sentiment does not reliably predict higher subsequent excess returns across firms or within firms over time. However, the directional consistency is noteworthy. The directional pattern is consistent with prior product-level evidence from Wu et al. (2020) and Jin (2023), which documents a positive association between review-based sentiment measures and subsequent stock returns. This suggests that company-level review sentiment may contain some information relevant to stock returns, even if that information does not translate into statistically significant return predictability under the current sample and data constraints. Furthermore, the high-coverage subsample provides suggestive cross-sectional evidence that customer sentiment may be more informative when it is based on a larger number of customer observations. This is broadly consistent with the wisdom of crowds argument advanced by Surowiecki (2004), where aggregated opinions may become more informative when they reflect broader underlying participation.

These findings contribute to the literature by suggesting that, in the present sample, company-level reviews do not provide the same clear return-predictive signal documented in the product-review literature. It instead suggests that the relationship between abnormal sentiment and returns does not transfer directly to the company-level review setting. One interpretation is that company-level reviews capture a broader and noisier set of customer experiences than product-level reviews. Unlike product-level reviews, which evaluate a specific good and may contain relatively precise information about product quality and demand (Chevalier and Mayzlin, 2006; Liu, 2006; Ludwig et al., 2013), company-level reviews aggregate consumer assessments across the full customer experience (Parasuraman et al., 1988; Richard and Allaway, 1993). While this breadth is an advantage for capturing firm-level customer experience, it may also introduce greater noise into the sentiment

signal. This interpretation is consistent with Zhu and Zhang (2010), who show that the impact of reviews varies substantially across contexts. Furthermore, the insignificant return estimates can be interpreted through the efficient market hypothesis (Fama, 1970). If customer reviews are publicly available, they should predict subsequent excess returns only if they contain value-relevant information that investors have not already incorporated into prices.

The lack of statistically significant return predictability is therefore consistent with two possible interpretations. First, the company-level sentiment signal may be too broad and noisy to generate a precise short-term return effect. Second, investors may incorporate this information quickly, leaving limited scope for predictable excess returns. Importantly, this does not imply that company-level reviews are irrelevant. The generally positive coefficient estimates, together with the pattern observed among high-coverage observations, point to a possible investor-relevant signal. However, in our sample, the signal appears too weak or noisy to yield precise estimates in the short-term return setting.

The second research question asks whether negative abnormal sentiment has stronger return relevance than positive abnormal sentiment. The results provide partial but cautious support for this prediction. Negative abnormal sentiment shows a more consistent relation to subsequent excess returns than positive abnormal sentiment, particularly in the cross-sectional analysis. However, the evidence does not support a strong general asymmetry, as the difference between positive and negative sentiment is not robust across specifications and is not clearly observed within firms over time. The strongest indication of asymmetric return relevance appears when negative sentiment is combined with high review dispersion. This suggests that negative customer sentiment may be more informative for returns when customer opinions are more inconsistent, although the evidence remains suggestive rather than conclusive.

The findings contribute to the literature by suggesting that the informational content of company-level reviews depends not only on the direction of customer sentiment, but also on the degree of disagreement across customer opinions. The more consistent relationship between negative sentiment and subsequent returns aligns with prior product-review evidence showing that negative reviews can have stronger economic effects than positive reviews (Wu et al., 2025; Tirunillai and Tellis, 2012). This interpretation is also consistent with prospect theory, which predicts that negative information receives greater weight

than positive information (Kahneman and Tversky, 1979). However, unlike the stronger asymmetry documented in some product-review settings, the evidence in this thesis is not robust across specifications and does not consistently appear within firms over time. One interpretation is that company-level reviews contain a broader and more heterogeneous set of customer experiences than product-level reviews (Parasuraman et al., 1988; Richard and Allaway, 1993). Negative sentiment may therefore capture many different types of negative experiences, including delivery problems, customer service issues, return processes, or product quality concerns, all of which may not have the same relevance for firm value. This heterogeneity may weaken the precision of the estimated asymmetric return effect.

At the same time, the dispersion result suggests that negative sentiment becomes more informative when customer opinions are more inconsistent. Higher dispersion may signal uncertainty about the firm's overall customer experience, thereby amplifying the return relevance of negative sentiment. This is consistent with Babić Rosario et al. (2016), who show that variability in product-reviews can increase perceived risk. This suggests that the financial relevance of negative company-level sentiment depends on the broader structure of the review signal, rather than sentiment direction alone. Taken together, the results suggest that negative sentiment may contain investor-relevant information, but that it does not produce a robust general asymmetry unless it is accompanied by higher review dispersion.

The third research question asks whether company-level abnormal customer sentiment predicts future revenue growth. The results provide the strongest support among the three research questions. Quarterly abnormal customer sentiment is positively associated with subsequent revenue growth, and the relation remains statistically significant after adding control variables. This suggests that company-level customer reviews contain information about future firm fundamentals. However, the result should not be interpreted as causal. The analysis shows that abnormal sentiment predicts future revenue growth, but it does not establish that changes in customer sentiment cause changes in revenue.

This finding contributes to the literature by showing that company-level reviews can predict future operating performance, even when their return-predictability is weaker. It extends prior research on product-level reviews and firm performance by showing that broader company-level reviews also contain information about future firm fundamentals. The findings position company-level reviews as a meaningful source of forward-looking

information about firm demand, consistent with Subrahmanyam and Titman’s (1999) concept of serendipitous information. The results further align with Jin (2023) and Huang (2018), who find that product-level reviews predict future revenue and earnings surprises. Finding this relationship in company-level Trustpilot reviews suggests that the revenue-predictive content of review sentiment extends beyond product-specific settings to broader firm-level evaluations.

Taken together, the strongest evidence emerges in the revenue growth analysis, where abnormal customer sentiment positively predicts subsequent revenue growth. By contrast, the return based analysis provides only limited evidence of return predictability, although they suggest that negative sentiment may become more informative when customer opinions are more dispersed. Therefore, the findings suggest that company-level customer reviews contain value-relevant information that is reflected more clearly in future firm fundamentals than in short-term stock price movements.

6.1 Implications

The findings have several implications for academic research. First, it extends the review literature from a product-level to a company-level setting, suggesting that broader customer experiences may also contain information relevant to firm outcomes. This extension is especially relevant for firm fundamentals, where abnormal company-level review sentiment predicts subsequent revenue growth, suggesting that such reviews capture demand-related information beyond product-specific evaluations. Second, the results further indicate that the informativeness of company-level review sentiment depends on contextual characteristics of the review signal. Review coverage, sentiment asymmetry, and review dispersion appear to influence the extent to which review sentiment relates to future firm performance, although to varying degrees. Future research should therefore account for these contextual factors rather than treating company-level review sentiment as a uniform signal.

The findings also carry implications for managers and investors. For managers, the revenue growth findings suggest that company-level reviews provide a market-based signal of customer experience that is associated with subsequent revenue growth. This underscores the importance of monitoring and responding to customer feedback, not only as a service quality measure, but also as a potential indicator of future demand

conditions. For investors and analysts, the results suggest that company-level review sentiment may provide a complementary source of information about firm prospects, particularly regarding future revenue development. However, the return-based evidence is directional rather than conclusive and should be interpreted alongside traditional financial measures rather than as a standalone signal.

6.2 Limitations and Future Research

Several limitations should be considered when interpreting the findings of this thesis. First, the sample is confined to Nordic listed retail firms with sufficient Trustpilot coverage over a relatively limited twelve-year period. This restricts the overall sample size and the generalizability of the findings. Nordic markets have distinct characteristics in terms of, for example, investor base and firm size, which may limit the extent to which the results apply to broader financial markets. The relatively small number of firms with sufficient review coverage may also reduce statistical power and partly explain the inconclusive results for the return-based tests. Future research could extend the analysis to larger samples across major financial markets over a longer time horizon to examine whether company-level review sentiment predicts stock returns more consistently across different market structures and investor environments.

Second, the reliance on Trustpilot and the focus on retail firms introduce limitations related to generalizability. As Yu et al. (2013) demonstrate, different media channels have heterogeneous effects on stock performance, meaning that the findings may reflect characteristics specific to Trustpilot rather than company-level review sentiment more broadly. In addition, Trustpilot allows firms to actively solicit reviews, which may introduce systematic bias in the sentiment distribution. The retail setting may further limit generalizability, since retail firms often have frequent and direct contact with end customers, making customer reviews more observable and more closely connected to consumer demand than in less customer-facing industries. Future research could replicate the analysis across other company-level review platforms and industries to examine whether the findings reflect a broader phenomenon or are specific to Trustpilot and the retail setting. Such an extension could also shed light on whether different types of company-level evaluations carry distinct informational content for financial markets.

Third, company-level review sentiment may include aspects irrelevant to returns,

contributing to noise. Unlike product-level reviews, company-level reviews capture a broader range of customer experiences, including service quality, delivery, customer support, post-purchase interactions, and overall brand perception (Parasuraman et al., 1988; Richard and Allaway, 1993). This broad scope is central to the contribution of the thesis, but may also make the sentiment signal less directly linked to short-term stock returns. Future research could therefore decompose company review text into specific topics or experiential dimensions, such as service quality, product quality, delivery, pricing, or customer support. This could help identify which aspects of the customer experience contain the most value-relevant information. A direct comparison between product-level and company-level reviews for the same firms would also help determine whether broader firm-level evaluations or more specific product-level evaluations provide stronger signals for stock returns and firm fundamentals.

Finally, the empirical design supports a predictive, but not causal, interpretation. While abnormal customer sentiment predicts future revenue growth, the analysis does not establish that changes in sentiment cause changes in revenue. The observed relation may instead be driven by underlying factors not fully captured by the control variables. Future research could therefore examine whether the documented relationship between customer sentiment and financial outcomes reflects a causal effect.

7. References

- Aksoy, L., Cooil, B., Groening, C. J., Keiningham, T. L., & Yalçın, A. (2008). The long-term stock market valuation of customer satisfaction. *Journal of Marketing*, 72(4), 105-122. <https://doi.org/10.1509/jmkg.72.4.105>
- Anderson, E. W. (1998). Customer satisfaction and word of mouth. *Journal of Service Research*, 1(1), 5-17. <https://doi.org/10.1177/109467059800100102>
- Archak, N., Ghose, A., & Ipeirotis, P. G. (2011). Deriving the pricing power of product features by mining consumer reviews. *Management Science*, 57(8), 1485-1509. <https://doi.org/10.1287/mnsc.1110.1370>
- Babić Rosario, A., Sotgiu, F., De Valck, K., & Bijmolt, T. H. A. (2016). The effect of electronic word of mouth on sales: A meta-analytic review of platform, product, and metric factors. *Journal of Marketing Research*, 53(3), 297-318. <https://doi.org/10.1509/jmr.14.0380>
- Barbieri, F., Espinosa Anke, L., & Camacho-Collados, J. (2022). XLM-T: Multilingual language models in Twitter for sentiment analysis and beyond. *arXiv*. <https://doi.org/10.48550/arXiv.2104.12250>
- Bell, D. R., Ledoit, O., & Wolf, M. (2014). A new portfolio formation approach to mispricing of marketing performance indicators: An application to customer satisfaction. *Customer Needs and Solutions*, 1(4), 263-276. <https://doi.org/10.1007/s40547-014-0028-6>
- Chakraborty, I., Kim, M., & Sudhir, K. (2022). Attribute sentiment scoring with online text reviews: Accounting for language structure and missing attributes. *Journal of Marketing Research*, 59(3), 510-531. <https://doi.org/10.1177/00222437211052500>
- Chen, H., De, P., Hu, Y. J., & Hwang, B. H. (2014). Wisdom of crowds: The value of stock opinions transmitted through social media. *Review of Financial Studies*, 27(5), 1367-1403. <https://doi.org/10.1093/rfs/hhu001>
- Chen, Y., & Xie, J. (2008). Online consumer review: Word-of-mouth as a new element

- of marketing communication mix. *Management Science*, 54(3), 477-491. <https://doi.org/10.1287/mnsc.1070.0810>
- Cheung, C. M. K., Lee, M. K. O., & Rabjohn, N. (2008). The impact of electronic word-of-mouth: The adoption of online opinions in online customer communities. *Internet Research*, 18(3), 229-247. <https://doi.org/10.1108/10662240810883290>
- Chevalier, J. A., & Mayzlin, D. (2006). The effect of word of mouth on sales: Online book reviews. *Journal of Marketing Research*, 43(3), 345-354. <https://doi.org/10.1509/jmkr.43.3.345>
- Cho, S., Pekgün, P., Janakiraman, R., & Wang, J. (2024). The competitive effects of online reviews on hotel demand. *Journal of Marketing*, 88(2), 40-60. <https://doi.org/10.1177/00222429231191449>
- Davoodi, L., Mezei, J., & Heikkilä, M. (2026). Aspect-based sentiment classification of user reviews to understand customer satisfaction of e-commerce platforms. *Electronic Commerce Research*, 26, 1417–1459. <https://doi.org/10.1007/s10660-025-09948-4>
- Fama, E. F. (1970). Efficient capital markets: A review of theory and empirical work. *The Journal of Finance*, 25(2), 383-417. <https://doi.org/10.2307/2325486>
- Fama, E. F., & MacBeth, J. D. (1973). Risk, return, and equilibrium: Empirical tests. *Journal of Political Economy*, 81(3), 607-636. <https://doi.org/10.1086/260061>
- Fornell, C., Mithas, S., Morgeson, F. V., & Krishnan, M. S. (2006). Customer satisfaction and stock prices: High returns, low risk. *Journal of Marketing*, 70(1), 3-14. <https://doi.org/10.1509/JMKG.2006.70.1.3>
- Fornell, C., Morgeson, F. V., III, & Hult, G. T. M. (2016). Stock returns on customer satisfaction do beat the market: Gauging the effect of a marketing intangible. *Journal of Marketing*, 80(5), 92-107. <https://doi.org/10.1509/jm.15.0229>
- Hennig-Thurau, T., Gwinner, K. P., & Gremler, D. D. (2002). Understanding relationship marketing outcomes: An integration of relational benefits and relationship quality. *Journal of Service Research*, 4(3), 230-247. <https://doi.org/10.1177/10946705020040>

- Huang, A. H. (2018). The customer knows best: The investment value of consumer opinions. *Journal of Financial Economics*, 128(1), 164-182. <https://doi.org/10.1016/j.jfineco.2018.02.001>
- Ittner, C. D., Larcker, D. F., & Taylor, D. J. (2009). Commentary - The stock market's pricing of customer satisfaction. *Marketing Science*, 28(5), 826-835. <https://doi.org/10.1287/mksc.1090.0526>
- Jacobson, R., & Mizik, N. (2009). The financial markets and customer satisfaction: Re-examining possible financial market mispricing of customer satisfaction. *Marketing Science*, 28(5), 810-819. <https://doi.org/10.1287/mksc.1090.0495>
- Jin, S. C. (2023). More than words: Can tone of consumer product reviews help predict firms' fundamentals? *Journal of Business Finance & Accounting*, 50(9-10), 1910-1942. <https://doi.org/10.1111/jbfa.12680>
- Kahneman, D., & Tversky, A. (1979). Prospect theory: An analysis of decision under risk. *Econometrica*, 47(2), 263-292. <https://doi.org/10.2307/1914185>
- Liu, Y. (2006). Word of mouth for movies: Its dynamics and impact on box office revenue. *Journal of Marketing*, 70(3), 74-89. <https://doi.org/10.1509/jmkg.70.3.074>
- Ludwig, S., de Ruyter, K., Friedman, M., Brüggem, E. C., Wetzels, M., & Pfann, G. (2013). More than words: The influence of affective content and linguistic style matches in online reviews on conversion rates. *Journal of Marketing*, 77(1), 87-103. <https://doi.org/10.1509/jm.11.0560>
- Luo, X., & Zhang, J. (2013). How do consumer buzz and traffic in social media marketing predict the value of the firm? *Journal of Management Information Systems*, 30(2), 213-238. <https://doi.org/10.2753/MIS0742-1222300208>
- Luo, X., Zhang, J., & Duan, W. (2013). Social media and firm equity value. *Information Systems Research*, 24(1), 146-163. <https://doi.org/10.1287/isre.1120.0462>
- Newey, W. K., & West, K. D. (1987). A simple, positive semi-definite, heteroskedasticity

- and autocorrelation consistent covariance matrix. *Econometrica*, 55(3), 703-708.
<https://doi.org/10.2307/1913610>
- Parasuraman, A., Zeithaml, V. A., & Berry, L. L. (1988). SERVQUAL: A multiple-item scale for measuring consumer perceptions of service quality. *Journal of Retailing*, 64(1), 12-40.
- Qiu, K., & Zhang, L. (2024). How online reviews affect purchase intention: A meta-analysis across contextual and cultural factors. *Data and Information Management*, 8, 100058. <https://doi.org/10.1016/j.dim.2023.100058>
- Raketa, D. (2024, July 11). How reviews and ratings affect clients' buying decisions. *Forbes*. <https://www.forbes.com/councils/forbesbusinessdevelopmentcouncil/2024/07/11/how-reviews-and-ratings-affect-clients-buying-decisions/>
- Richard, M. D., & Allaway, A. W. (1993). Service quality attributes and choice behavior. *Journal of Services Marketing*, 7(1), 59-68. <https://doi.org/10.1108/08876049310026105>
- Subrahmanyam, A., & Titman, S. (1999). The going-public decision and the development of financial markets. *The Journal of Finance*, 54(3), 1045-1082. <https://doi.org/10.1111/0022-1082.00136>
- Surowiecki, J. (2004). *The wisdom of crowds: Why the many are smarter than the few and how collective wisdom shapes business, economies, societies, and nations*. Doubleday.
- Tirunillai, S., & Tellis, G. J. (2012). Does chatter really matter? Dynamics of user-generated content and stock performance. *Marketing Science*, 31(2), 198-215. <https://doi.org/10.1287/mksc.1110.0682>
- Wu, D., Zhou, Z., & Su, Z. (2025). Consumer negative opinions on stock returns: Evidence from e-commerce reviews in China. *Financial Innovation*, 11(1), 71. <https://doi.org/10.1186/s40854-024-00746-9>
- Wu, J., Xu, K., & Zhao, J. (2020). Predicting long-term returns of individual stocks with online reviews. *Neurocomputing*, 417, 406-418. <https://doi.org/10.1016/j.neucom.202>

- Yu, Y., Duan, W., & Cao, Q. (2013). The impact of social and conventional media on firm equity value: A sentiment analysis approach. *Decision support systems*, 55(4), 919-926. <https://doi.org/10.1016/j.dss.2012.12.028>
- Zhu, F., & Zhang, X. (2010). Impact of online consumer reviews on sales: The moderating role of product and consumer characteristics. *Journal of Marketing*, 74(2), 133-148. <https://doi.org/10.1509/jm.74.2.133>

8. Appendix

Table A.1
Fixed-Effects Panel Regressions of Excess Returns by Review Coverage

This table presents the results from panel regressions of next-month excess returns on abnormal review sentiment and control variables with firm and month fixed-effects, separately for low- and high-coverage observations. Columns (i) - (iv) report low-coverage observations, while columns (v) - (viii) report high-coverage observations. Within each group, the specifications progressively add controls: no controls, review-related controls, standard firm-level controls, and extended firm-level controls. Review-related controls include abnormal review count and review dispersion. Standard firm-level controls include market capitalization, book-to-market ratio, and momentum. Extended controls further include leverage and gross profitability. Firm-clustered standard errors are reported in parentheses.

	Low-Coverage				High-Coverage			
	(i) No controls	(ii) Review controls	(iii) Standard controls	(iv) Extended controls	(v) No controls	(vi) Review controls	(vii) Standard controls	(viii) Extended controls
Abnormal review sentiment	0.0047 (0.0164)	0.0013 (0.0188)	0.0022 (0.0190)	0.0016 (0.0198)	0.0125 (0.0165)	0.0145 (0.0167)	0.0113 (0.0159)	0.0113 (0.0158)
Abnormal review count		-0.0002 (0.0001)	-0.0002 (0.0001)	-0.0002 (0.0002)		$>-0.0001^{36}$ $(<0.0001)^{39}$	$>-0.0001^{37}$ $(<0.0001)^{40}$	$>-0.0001^{38}$ $(<0.0001)^{41}$
Review dispersion		0.0217 (0.0353)	0.0289 (0.0334)	0.0259 (0.0342)		-0.0459 (0.0324)	-0.0406 (0.0294)	-0.0396 (0.0305)
Log MCAP			-0.0381 (0.0246)	-0.0393 (0.0247)			-0.0497*** (0.0100)	-0.0528*** (0.0105)
Book-to-market			0.0260 (0.0236)	0.0255 (0.0237)			-0.0056 (0.0083)	-0.0082 (0.0079)
Momentum			0.0048 (0.0140)	0.0033 (0.0148)			0.0167** (0.0068)	0.0168** (0.0069)
Leverage				0.0016 (0.0921)				-0.0542 (0.0401)
Gross profitability				0.1723 (0.1236)				-0.0419 (0.0334)
Months	63	63	63	63	131	131	130	130
Avg. firms per month	48	48	48	48	58	58	57	57
Adjusted R ²	0.152	0.152	0.169	0.169	0.209	0.210	0.244	0.246

, **, and * indicate statistical significance at the 10%, 5%, and 1% levels, respectively.*

³⁶ -5.84×10^{-6} ³⁷ -5.51×10^{-6} ³⁸ -5.12×10^{-6} ³⁹ 5.39×10^{-6} ⁴⁰ 6.10×10^{-6} ⁴¹ 5.98×10^{-6}

Table A.2
Fixed-Effects Panel Regressions on Asymmetric Sentiment and Excess Returns

This table presents the results from panel regressions of next-month excess returns on positive and negative abnormal review sentiment and control variables with firm and month fixed-effects. Column (i) presents the baseline specification without additional controls. Column (ii) includes review-related controls, specifically abnormal review count and dispersion. Column (iii) adds standard firm-level controls, including market capitalization, book-to-market ratio, and momentum. Column (iv) further extends the specification by including leverage and gross profitability. $\beta_{\text{pos}} - \beta_{\text{neg}}$ reports the difference between the positive and negative sentiment coefficients. Firm-clustered standard errors are reported in parentheses.

	(i) No controls	(ii) Review controls	(iii) Standard controls	(iv) Extended controls
Positive abnormal sentiment	0.0038 (0.0231)	0.0047 (0.0232)	-0.0070 (0.0224)	-0.0060 (0.0225)
Negative abnormal sentiment	-0.0125 (0.0247)	-0.0152 (0.0237)	-0.0217 (0.0205)	-0.0184 (0.0198)
$\beta_{\text{pos}} - \beta_{\text{neg}}$	0.0163 (0.0230)	0.0199 (0.0226)	0.0148 (0.0225)	0.0124 (0.0224)
Abnormal review count		$>-0.0001^{42}$ (<0.0001) ⁴⁵	$>-0.0001^{43}$ (<0.0001) ⁴⁶	$>-0.0001^{44}$ (<0.0001) ⁴⁷
Review dispersion		-0.0096 (0.0200)	-0.0075 (0.0193)	-0.0076 (0.0196)
Log MCAP			-0.0469*** (0.0088)	-0.0507*** (0.0094)
Book-to-market			-0.0029 (0.0090)	-0.0059 (0.0085)
Momentum			0.0145** (0.0060)	0.0145** (0.0061)
Leverage				-0.0619* (0.0360)
Gross profitability				-0.0389 (0.0303)
Months	131	131	130	130
Avg. firms per month	68	68	67	67
Adjusted R ²	0.198	0.198	0.229	0.231

, **, and * indicate statistical significance at the 10%, 5%, and 1% levels, respectively.*

⁴² -5.02×10^{-6} ⁴³ -3.96×10^{-6} ⁴⁴ -3.40×10^{-6} ⁴⁵ 5.18×10^{-6} ⁴⁶ 6.00×10^{-6} ⁴⁷ 5.88×10^{-6}

Table A.3
Fixed-Effects Panel Regressions on Negative Sentiment, Review Dispersion, and Subsequent Excess Returns

This table presents results from panel regressions of next-month excess returns on abnormal review sentiment and its interaction with dispersion. Review dispersion is standardized within month before constructing the interaction term, so zero corresponds to the monthly mean. Accordingly, the coefficient on negative abnormal sentiment magnitude is evaluated at mean dispersion, while the coefficient on review dispersion is evaluated when negative abnormal sentiment magnitude equals zero. Column (i) presents the baseline specification without additional controls. Column (ii) includes review-related controls. Column (iii) adds standard firm-level controls, including market capitalization, book-to-market ratio, and momentum. Column (iv) further extends the specification by including leverage and gross profitability. Firm-clustered standard errors are reported in parentheses.

	(i) No controls	(ii) Review controls	(iii) Standard controls	(iv) Extended controls
Negative abnormal sentiment	-0.0152 (0.0245)	-0.0163 (0.0243)	-0.0234 (0.0210)	-0.0201 (0.0204)
Positive abnormal sentiment	0.0029 (0.0231)	0.0042 (0.0233)	-0.0077 (0.0226)	-0.0068 (0.0227)
Review dispersion	-0.0001 (0.0039)	-0.0001 (0.0039)	0.0011 (0.0039)	0.0014 (0.0042)
Negative sentiment $\times$ dispersion	-0.0177 (0.0183)	-0.0178 (0.0182)	-0.0249 (0.0183)	-0.0269 (0.0189)
Abnormal review count		$>-0.0001^{48}$ (<0.0001) ⁵¹	$>-0.0001^{49}$ (<0.0001) ⁵²	$>-0.0001^{50}$ (<0.0001) ⁵³
Log MCAP			-0.0469*** (0.0089)	-0.0508*** (0.0095)
Book-to-market			-0.0028 (0.0091)	-0.0057 (0.0085)
Momentum			0.0146** (0.0061)	0.0146** (0.0062)
Leverage				-0.0636* (0.0370)
Gross profitability				-0.0383 (0.0307)
Months	131	131	130	130
Avg. firms per month	68	68	67	67
Adjusted R ²	0.198	0.198	0.229	0.231

, **, and * indicate statistical significance at the 10%, 5%, and 1% levels, respectively.*

⁴⁸ -5.05×10^{-6} ⁴⁹ -3.98×10^{-6} ⁵⁰ -3.42×10^{-6} ⁵¹ 5.17×10^{-6} ⁵² 5.97×10^{-6} ⁵³ 5.84×10^{-6}

Table A.4
Robustness - Star Rating

This table replicates all main analyses using the abnormal review rating as the main explanatory variable instead of abnormal review sentiment. All sample construction, thresholds, controls, and specifications are otherwise identical to the main analysis. Each row block reports the coefficient on the main variable, the associated standard error in parentheses, average firms per month, and the number of months used in estimation across specifications (i) - (viii). Columns (i) - (iv) report Fama-MacBeth results, while columns (v) - (viii) report fixed-effects results. Within each group, the specifications progressively add controls following the main analysis. For the Fama-MacBeth regressions, parenthesized values are Newey-West standard errors and avg. firms reports the average number of firms per monthly cross-section. For the fixed-effects regressions, which include firm and month fixed-effects, parenthesized values are firm-clustered standard errors and avg. firms reports unique firms in the sample. Quarterly rating (RQ3) is estimated with fixed-effects only and reports the number of quarters in place of months. Sign changes relative to the main analysis occur only in statistically insignificant estimates, suggesting imprecision rather than evidence of an opposite relation.

		Fama-MacBeth				Fixed-Effects			
		(i)	(ii)	(iii)	(iv)	(v)	(vi)	(vii)	(viii)
RQ1	Abnormal review rating	0.0076	0.0073	0.0073	0.0055	0.0035	0.0042	0.0029	0.0027
	SE	(0.0049)	(0.0051)	(0.0058)	(0.0058)	(0.0039)	(0.0039)	(0.0040)	(0.0040)
	Avg. firms per month / firms	39.81	39.81	38.66	38.23	68	68	67	67
	N months	77	77	77	77	131	131	130	130
	High coverage rating	0.0097*	0.0107*	0.0090	0.0079	0.0030	0.0038	0.0029	0.0030
	SE	(0.0058)	(0.0060)	(0.0074)	(0.0073)	(0.0051)	(0.0053)	(0.0051)	(0.0051)
	Avg. firms per month / firms	29.22	29.22	28.17	27.92	58	58	57	57
	N months	77	77	77	77	131	131	130	130
	Low coverage rating	-0.0017	0.0012	0.0033	0.0039	0.0093	0.0092	0.0086	0.0087
	SE	(0.0080)	(0.0077)	(0.0090)	(0.0115)	(0.0067)	(0.0078)	(0.0079)	(0.0081)
	Avg. firms per month / firms	18.90	18.90	18.76	18.28	48	48	48	48
	N months	77	77	77	77	131	131	130	130
RQ2	Positive abnormal rating	0.0077	0.0121	0.0068	0.0033	-0.0009	-0.0003	-0.0036	-0.0032
	SE	(0.0095)	(0.0104)	(0.0090)	(0.0088)	(0.0062)	(0.0063)	(0.0059)	(0.0059)
	Avg. firms per month / firms	39.81	39.81	38.66	38.23	68	68	67	67
	N months	77	77	77	77	131	131	130	130
	Negative abnormal rating	-0.0112	-0.0055	-0.0142*	-0.0131	-0.0080	-0.0089	-0.0098	-0.0091
	SE	(0.0075)	(0.0069)	(0.0085)	(0.0113)	(0.0083)	(0.0082)	(0.0075)	(0.0074)
	Avg. firms per month / firms	39.81	39.81	38.66	38.23	68	68	67	67
	N months	77	77	77	77	131	131	130	130
	$\beta_{\text{pos}} - \beta_{\text{neg}}$	0.0190*	0.0176*	0.0211*	0.0164	0.0071	0.0086	0.0062	0.0059
	SE	(0.0100)	(0.0095)	(0.0123)	(0.0134)	(0.0078)	(0.0078)	(0.0080)	(0.0080)
	Avg. firms per month / firms	39.81	39.81	38.66	38.23	68	68	67	67
	N months	77	77	77	77	131	131	130	130
RQ3	Negative rating $\times$ dispersion	-0.0126	0.0034	-0.0242	-0.0432	-0.0023	-0.0023	-0.0057	-0.0069
	SE	(0.0267)	(0.0339)	(0.0350)	(0.0313)	(0.0070)	(0.0069)	(0.0072)	(0.0074)
	Avg. firms per month / firms	39.81	39.81	38.66	38.23	68	68	67	67
	N months	77	77	77	77	131	131	130	130
	Quarterly abnormal rating	—	—	—	—	0.0076	0.0213	0.0221	0.0238
	SE	—	—	—	—	(0.0147)	(0.0167)	(0.0168)	(0.0162)
	Avg. firms per quarter/ firms	—	—	—	—	66	66	66	66
	N quarters	—	—	—	—	43	43	43	43

*, **, and *** indicate statistical significance at the 10%, 5%, and 1% levels, respectively.

Table A.5
Robustness - 10-Review Minimum Threshold

This table replicates all main analyses using a minimum threshold of 10 reviews per month instead of the main analysis threshold of 5. Each row block reports the coefficient on the main variable, the associated standard error in parentheses, average firms per month, and the number of months used in estimation across specifications (i) - (viii). Columns (i) - (iv) report Fama-MacBeth results, while columns (v) - (viii) report fixed-effects results. Within each group, the specifications progressively add controls following the main analysis. For the Fama-MacBeth regressions, parenthesized standard errors are Newey-West adjusted and avg. firms reports the average number of firms per monthly cross-section. For the fixed-effects regressions, which include firm and month fixed-effects, parenthesized standard errors are firm-clustered and avg. firms reports unique firms in the sample. Low coverage Fama-MacBeth could not be estimated due to insufficient sample size under the 10-review threshold. Quarterly sentiment (RQ3) is estimated with fixed-effects only and reports the number of quarters in place of months. Sign changes relative to the main analysis occur only in statistically insignificant estimates, suggesting imprecision rather than evidence of an opposite relation.

		Fama-MacBeth				Fixed-Effects			
		(i)	(ii)	(iii)	(iv)	(v)	(vi)	(vii)	(viii)
RQ1	Abnormal review sentiment	0.0231	0.0294*	0.0207	0.0248	0.0153	0.0178	0.0137	0.0118
	SE	(0.0145)	(0.0153)	(0.0168)	(0.0213)	(0.0147)	(0.0146)	(0.0143)	(0.0142)
	Avg. firms per month / firms	36.24	36.24	34.90	34.46	57	57	56	56
	N months	63	63	63	63	131	131	130	130
	High coverage sentiment	0.0251	0.0305	0.0205	0.0292	0.0179	0.0192	0.0182	0.0176
	SE	(0.0184)	(0.0187)	(0.0210)	(0.0244)	(0.0171)	(0.0174)	(0.0164)	(0.0166)
	Avg. firms per month / firms	30.86	30.86	29.60	29.30	54	54	53	53
	N months	63	63	63	63	131	131	130	130
	Low coverage sentiment	—	—	—	—	0.0138	0.0061	0.0112	0.0029
	SE	—	—	—	—	(0.0357)	(0.0289)	(0.0300)	(0.0299)
	Avg. firms per month / firms	—	—	—	—	34	34	34	34
	N months	—	—	—	—	55	55	55	55
RQ2	Positive abnormal sentiment	0.0322	0.0343	-0.0022	0.0183	0.0138	0.0147	-0.0028	-0.0017
	SE	(0.0254)	(0.0260)	(0.0307)	(0.0353)	(0.0264)	(0.0257)	(0.0274)	(0.0271)
	Avg. firms per month / firms	36.24	36.24	34.90	34.46	57	57	56	56
	N months	63	63	63	63	131	131	130	130
	Negative abnormal sentiment	-0.0079	-0.0239	-0.0265	0.0157	-0.0169	-0.0210	-0.0317	-0.0266
	SE	(0.0246)	(0.0226)	(0.0307)	(0.0449)	(0.0407)	(0.0400)	(0.0358)	(0.0355)
	Avg. firms per month / firms	36.24	36.24	34.90	34.46	57	57	56	56
	N months	63	63	63	63	131	131	130	130
	$\beta_{\text{pos}} - \beta_{\text{neg}}$	0.0401	0.0582**	0.0244	0.0025	0.0306	0.0358	0.0288	0.0249
	SE	(0.0276)	(0.0274)	(0.0310)	(0.0359)	(0.0305)	(0.0302)	(0.0289)	(0.0289)
	Avg. firms per month / firms	36.24	36.24	34.90	34.46	57	57	56	56
	N months	63	63	63	63	131	131	130	130
RQ3	Negative sentiment $\times$ dispersion	-0.0907	-0.1127	-0.1250	-0.1670	-0.0203	-0.0195	-0.0329	-0.0373
	SE	(0.0630)	(0.0689)	(0.0751)	(0.1220)	(0.0293)	(0.0289)	(0.0285)	(0.0294)
	Avg. firms per month / firms	36.24	36.24	34.90	34.46	57	57	56	56
	N months	63	63	63	63	131	131	130	130
	Quarterly abnormal sentiment	—	—	—	—	0.0574	0.1070**	0.1111**	0.1053**
RQ3	SE	—	—	—	—	(0.0436)	(0.0435)	(0.0460)	(0.0439)
	Avg. firms per quarter / firms	—	—	—	—	56	56	56	56
	N quarters	—	—	—	—	43	43	43	43

*, **, and *** indicate statistical significance at the 10%, 5%, and 1% levels, respectively.

AI Appendix

Generative artificial intelligence tools, including ChatGPT (OpenAI), Claude (Anthropic), and Grammarly were used as supportive tools throughout the thesis process. These tools were primarily used for proofreading, improving grammar and sentence structure, and enhancing readability and flow. They also assisted in formatting tasks such as creating reference lists, creating LaTeX table code, and supporting Python coding and debugging throughout the code generation phase. AI tools were also used to ask clarifying questions and critically evaluate drafts and analytical reasoning. Nonetheless, all analysis, interpretations, conclusions and research decisions are our own, developed based on our own understanding, data analysis and review of academic literature. AI tools were used only to support and refine our work, not to replace our own reasoning.

We identified several risks associated with the use of generative AI, particularly the risk of hallucinated information, inaccurate outputs and unreliable source recommendation. To manage these risks, we did not rely on AI-generated references or factual claims, rather only used our own identified sources verified by ourselves. No output were interpreted as factual and therefore were carefully reviewed, edited and verified before inclusion in the thesis.

Throughout the process, we learned that generative AI can serve as an effective complement for improving efficiency, coding support and enhancing the clarity of the text. However, we also found that such tools cannot replace independent critical thinking, academic judgment or subject knowledge and that responsible use requires continuous verification and oversight.