



*Does investor sentiment affect the
cross-section of stock returns in the
post-2000 period?*

Jonas Porthun

Dissertation written under the supervision of Professor Mário Meira

Dissertation submitted in partial fulfilment of requirements for the MSc in
Finance, at the Universidade Católica Portuguesa, 04.01.2026.

ABSTRACT

This master's thesis examines the influence of investor sentiment on cross-sectional stock returns in the period from 2000 to 2023. While classical literature demonstrates sentiment-driven mispricing primarily in data from before the mid-2000s, this study offers an updated analysis for a period characterised by structural market changes and multiple crises.

Based on the sentiment index developed by Baker and Wurgler, the study aims to identify which company characteristics make firms particularly sensitive to market sentiment. Monthly return data from CRSP and quarterly data from Compustat for U.S. companies are used for empirical testing. The methodology involves the creation of portfolios based on eight company characteristics such as size, age, volatility and profitability. In addition, regression models are applied, controlling for the risk factors identified by Fama and French as well as momentum.

The key findings show that investor sentiment continues to affect cross-sectional stock returns, although those effects are weaker than in earlier studies. After controlling for risk factors and fixed effects, firm size remains the most robustly sentiment-sensitive characteristic in portfolio-level analysis, while firm-level regressions show persistent sentiment interactions with both size and book-to-market ratio. Other speculative characteristics lose statistical significance in modern market environments. This paper thus provides up-to-date evidence of which effects persist in today's financial markets, which are characterised by automated trading and changed access for private investors.

Title: Does investor sentiment affect the cross-section of stock returns in the post-2000 period?

Author: Jonas Porthun

Keywords: Investor sentiment, cross-sectional stock returns, behavioural finance, asset pricing, firm characteristics, Baker and Wurgler sentiment index, Fama–French factors, U.S. equity market

RESUMO

Esta tese de mestrado examina a influência do sentimento dos investidores nos retornos transversais das ações entre 2000 e 2023. Enquanto a literatura clássica documenta erros de precificação induzidos pelo sentimento sobretudo antes de meados dos anos 2000, este estudo apresenta uma análise atualizada para um período marcado por mudanças estruturais nos mercados financeiros e por múltiplas crises.

Com base no índice de sentimento desenvolvido por Baker e Wurgler, o estudo identifica quais características empresariais tornam as empresas particularmente sensíveis ao sentimento do mercado. A análise empírica utiliza dados mensais de rendimentos da CRSP e dados trimestrais da Compustat para empresas norte-americanas. A metodologia inclui a construção de carteiras com base em oito características das empresas, como dimensão, idade, volatilidade e rentabilidade, bem como modelos de regressão que controlam os fatores de risco de Fama e French e o fator momentum.

Os resultados mostram que o sentimento dos investidores continua a afetar os retornos transversais das ações, embora esses efeitos sejam mais fracos do que nos estudos anteriores. Após o controlo dos fatores de risco e dos efeitos fixos, a dimensão da empresa permanece a característica mais sensível ao sentimento na análise ao nível das carteiras, enquanto as regressões ao nível da empresa revelam interações persistentes entre o sentimento, a dimensão e o rácio book-to-market. Outras características especulativas perdem significância estatística em ambientes de mercado modernos, caracterizados por maior automatização da negociação e alterações no acesso dos investidores de retalho.

Título: O sentimento dos investidores afeta a secção transversal dos retornos das ações no período pós-2000?

Autor: Jonas Porthun

Palavras-chave: Sentimento dos investidores, retornos transversais das ações, finanças comportamentais, avaliação de ativos, características das empresas, índice de sentimento de Baker e Wurgler, fatores de Fama–French, mercado acionista dos Estados Unidos

TABLE OF CONTENTS

ABSTRACT	II
RESUMO	III
LIST OF TABLES.....	VI
LIST OF FIGURES.....	VI
1. INTRODUCTION.....	1
1.1 MOTIVATION AND RESEARCH CONTEXT	1
1.2 RESEARCH QUESTIONS.....	2
1.3 CONTRIBUTION AND KEY FINDINGS.....	2
1.4 STRUCTURE OF THE THESIS	3
2. LITERATURE REVIEW	3
2.1 ASSET PRICING AND THE CROSS-SECTION OF RETURNS.....	3
2.1.1 RISK-BASED EXPLANATIONS	3
2.1.2 BEHAVIOURAL FINANCE PERSPECTIVES	4
2.2 DEFINITION OF INVESTOR SENTIMENT	5
2.3 SENTIMENT AND CROSS-SECTIONAL RETURNS.....	5
2.3.1 U.S. EVIDENCE.....	5
2.3.2 INTERNATIONAL EVIDENCE	6
2.3.3 CONTRARIAN FINDINGS	6
3. DATA AND METHODOLOGY	7
3.1 DATA SOURCES AND SAMPLE CONSTRUCTION.....	7
3.1.1 SENTIMENT DATA	8
3.1.2 RETURN DATA.....	11
3.1.3 ACCOUNTING DATA	12
3.2 DESCRIPTIVE STATISTICS	13
3.3 EMPIRICAL METHODOLOGY	15
3.3.1 SAMPLE PREPARATION AND PORTFOLIO CONSTRUCTION	15
3.3.2 TIME-SERIES SPREAD REGRESSION	16
3.3.3 FIRM-LEVEL PANEL REGRESSION	17
4. EMPIRICAL RESULTS	18
4.1 DESCRIPTIVE PORTFOLIO EVIDENCE (UNCONDITIONAL).....	18
4.2 PREDICTIVE SPREAD REGRESSION (PORTFOLIO LEVEL)	20
4.3 FIRM-LEVEL EVIDENCE (PANEL REGRESSIONS)	23

4.4 SYNTHESIS OF PORTFOLIO-LEVEL AND FIRM-LEVEL EVIDENCE	26
5. DISCUSSION	26
5.1 INTERPRETATION OF MAIN FINDINGS	27
5.2 COMPARISON WITH BAKER AND WURGLER (2006)	28
5.3 SENTIMENT IN THE FACE OF MARKET EVOLUTION AND CRISES	29
5.4 PRACTICAL IMPLICATIONS	29
5.5 LIMITATIONS AND FUTURE RESEARCH	29
5.5.1 SAMPLE AND DATA LIMITATIONS.....	29
5.5.2 METHODOLOGICAL CONSTRAINTS	30
5.6 DIRECTIONS FOR FUTURE RESEARCH.....	30
6. CONCLUSION	31
REFERENCES.....	VII
APPENDIX I: DATA CONSTRUCTION.....	X
APPENDIX II: PORTFOLIO CONSTRUCTION DETAILS	XIX
APPENDIX III: ROBUSTNESS CHECKS.....	XX

LIST OF TABLES

Table 1: Summary statistics of the dataset after data cleaning.....	12
Table 2: Summary statistics sorted by sentiment _{t-1}	13
Table 3: Summary of characteristic spreads by sentiment regime.....	17
Table 4: Univariate Regression of monthly characteristic spreads.....	19
Table 5: Multivariate Regression of monthly characteristic spreads.....	20
Table 6: Firm-level panel regression with one month lagged sentiment.....	23
Table A1: Summary Statistics of the Merged Dataset (Before Cleaning).....	XVII
Table A2: Summary Statistics of the Merged Dataset by Sentiment (Before Cleaning)....	XVIII
Table A3: Portfolio Decile Returns – Excerpt from Variable Size.....	XIX
Table A4: Univariate Regression of characteristic spreads (sentiment dummy).....	XX
Table A5: Multivariate Regression of characteristic spreads (sentiment dummy).....	XXI
Table A6: Firm-Level Panel Regressions with Bootstrap Standard Errors.....	XXII

LIST OF FIGURES

Figure 1: Standardized investor sentiment index (2000-2023).....	9
--	---

1. INTRODUCTION

The introduction describes the motivation for this study and provides the research context. It then presents the research questions, summarizes the central findings, and shows the structure of the thesis.

1.1 MOTIVATION AND RESEARCH CONTEXT

Sentiment is a widely used concept in both academic research and financial media. Terms such as “market sentiment” or “investor sentiment” are frequently used to explain positive as well as negative stock market movements, when there are no fundamental factors to attribute to. This study aims to identify which firm characteristics make companies more prone to changes in investor sentiment and whether certain types of firms systematically benefit or suffer during periods of high or low investor sentiment.

To analyse sentiment empirically, investor sentiment must be measurable and quantifiable. Baker and Wurgler’s (2006) sentiment index therefore serve as the foundation of this empirical analysis. Firm reactions are examined using stock market returns. From a practical perspective, this study offers relevant insights, as investor sentiment has become increasingly important in recent decades. Financial markets have become more accessible to retail investors through digital brokerage platforms, while periods of low interest rates have increased the need to allocate savings towards equity markets. With the growing community of investors, financial markets continue to follow behavioural patterns, although the impact and persistence of these effects may have evolved due to structural changes in market composition and trading technology. Social media platforms like X and Reddit, where retail investors exchange and discuss ideas can boost those movements.

In the literature the cross-section of returns is a topic discussed from many different angles and perspectives. Early research focused on risk-based explanations, most popular the three-factor model proposed by Fama and French (1993), which explains equity returns using market, size and value factors. Further studies expanded this framework by introducing additional factors, including momentum, using past returns to predict future performance. With the rise of behavioural finance, spanning from everyday purchasing decisions at the grocery store to investing decisions at the stock markets, this topic was tested against the cross-section of returns as well. The irrationality of the investors mind was found to be the reason for stock market

anomalies, like bubbles, overreactions, underreactions and crashes. Within this behavioural framework, Baker and Wurgler (2006) build on insights from behavioural finance and defined investor sentiment as a belief about future cash flows and risks, which can't be quantified. They show that sentiment can systematically affect the cross-section of stock returns, particularly for firms that are difficult to value or arbitrage. Their findings have been supported for different markets and regions, however most of this literature uses data ending before the mid-2000s. This study provides a renewed assessment of the sentiment driven cross-section of returns. The data used spans from 2000-2023. This period is characterised by major structural changes and multiple financial crises. Within this time investors experienced the financial crisis 2008, the COVID-19 crash with a rapid recovery and geopolitical tensions, such as Ukraine conflict. In addition, automated trading systems gained increasing share in trading volumes. Those developments call for a revision of the existing literature.

1.2 RESEARCH QUESTIONS

As the baseline for the methodology and conceptual procedure serves the paper from Baker and Wurgler (2006). To ensure that the thesis follows a clear and consistent structure, the following research questions are formulated:

- (1) Does investor sentiment negatively predict future stock returns?
- (2) Do cross-sectional return premia vary with investor sentiment?
- (3) In high sentiment phases, do firms with characteristics typically associated with speculation (young age, high volatility, and low profitability) exhibit stronger sensitivity to sentiment-driven mispricing?
- (4) Does the predictive power of sentiment differ across firm characteristics and weaken once firm and industry fixed effects are included?

1.3 CONTRIBUTION AND KEY FINDINGS

Stock return data are obtained from CRSP, while accounting data are retrieved from Compustat via WRDS. After cleaning and merging the dataset, firms are sorted into decile portfolios based on firm characteristics, following Baker and Wurgler (2006). Relative to Baker and Wurgler (2006), this thesis (1) updates the sample to 2000–2023, (2) aligns monthly returns to quarterly accounting information using post-earnings months, (3) combines portfolio spread regressions

and firm-level fixed-effects panel regressions, and (5) incorporates an alternative MSCI-style momentum measure as a robustness check.

The central results of this work confirm that sentiment affects the cross-section of returns. After controlling for risk factors and fixed effects, firm size remains the most robustly sentiment-sensitive characteristic at the portfolio level, while firm-level regressions reveal persistent sentiment interactions with both firm size and book-to-market ratio. These findings provide updated evidence on sentiment effects in modern financial markets and clarify which sentiment-related effects persist when combining portfolio-level and firm-level empirical approaches.

1.4 STRUCTURE OF THE THESIS

The remainder of this dissertation is structured as follows. Existing literature is reviewed in Section 2. Section 3 describes the data and methodology, the empirical results are presented in Section 4 and Section 5 discusses the findings and concludes.

2. LITERATURE REVIEW

The following section covers existing findings about the relationship between cross-sectional returns and investor sentiment. The relationship between returns and various firm characteristics has long been of great interest to finance academics.

2.1 ASSET PRICING AND THE CROSS-SECTION OF RETURNS

2.1.1 RISK-BASED EXPLANATIONS

Fama and French (1993), pioneers in this field of research, proposed a three-factor model including market, size, and value factors to explain equity returns. Their results show that small-cap stocks and high book-to-market (value) stocks recorded higher average returns Proving that these factors capture systematic risk not explained by the classical capital asset pricing model. In a later study, Fama and French (1996) expand their three-factor model into a multi-factor approach, including earnings-price ratio and leverage. The results highlighted the relationship between firm-specific characteristics and asset pricing. Until then, existing studies focus on company-specific characteristics, ignoring returns from the past. Meanwhile, Jegadeesh and Titman (1993) analyse the momentum effect, finding that stocks performing well over the past

3 to 12 months continue to outperform in the following 3 to 12 months. Their findings validate the hypothesis that momentum is not explained by risk or the Fama-French factor models. Other researchers found evidence of additional factors that influence excess returns and are not explained by traditional risk-based models. Lakonishok et al. (1994) study value strategies and observe long-term excess positive returns compared to growth strategies, stemming from buying low price-to-book stocks. They argue that investors assign high importance to past returns, leading to asset mispricing. Their work suggests that behavioural biases, such as investor overreaction, are driving returns rather than risk alone. Daniel and Titman (1997) further question traditional risk-based models, finding that cross-sectional returns are better explained by firm characteristics than systematic risk factors.

2.1.2 BEHAVIOURAL FINANCE PERSPECTIVES

As the classical capital asset pricing theory expanded to include more factors, the importance of behavioural finance theory became clear. Richard Thaler (1985) laid the foundation for modern behavioural finance. Although his work focused more on everyday financial decisions than asset pricing, it offered key insights into systematic decision-making bias. His popular paper introduced mental accounting, showing how systematic and predictable irrational financial behaviour is, including in investment decisions and asset pricing. In his following career, Thaler published studies and books on the psychology of finance, introducing concepts like the theory of self-control, which explains why individuals often act against their best interests (Thaler and Shefrin 1981). In his conceptual work, *Advances in Behavioural Finance*, he linked investor behaviour to asset pricing anomalies such as bubbles, overreactions, underreactions, momentum, calendar effects, and crashes (Thaler 1993).

Barberis and Thaler (2003) assess investor sentiment, providing an overview of behavioural finance literature, summarizing both theory and empirical evidence. They identify investor sentiment as a key driver of asset pricing anomalies and market inefficiencies, driven by psychological biases and herd behaviour rather than rational fundamentals. They note that sentiment-driven effects are particularly visible when arbitrage is limited and costly, and periods of high sentiment predict lower subsequent returns, especially for hard-to-value stocks.

2.2 DEFINITION OF INVESTOR SENTIMENT

Various methods have emerged to measure and define investor sentiment. Baker and Wurgler (2006) define investor sentiment as a “soft” belief about future cash flows and investment risk, not reasoned by “hard” facts, reflecting subjective optimism or pessimism. Their composite sentiment index includes six market-based variables: closed-end fund discount, share turnover, number and first-day returns on IPOs, equity share in new issues, dividend premium, and net mutual fund flows. In a 2007 review, they broaden the definition, describing sentiment as irrational optimism or pessimism that collectively moves the market, distinguishing between “broad-based” sentiment and sentiment toward specific assets (Baker and Wurgler 2006; Baker and Wurgler 2007).

2.3 SENTIMENT AND CROSS-SECTIONAL RETURNS

Barberis et al. (1998) describe investor sentiment as irrational beliefs or future expectations about asset values, driven by psychological biases. Their model discusses two biases: conservatism, representing underreaction to news, and representativeness, indicating overreaction to recurring patterns. Further studies support and refine this definition. Stambaugh et al. (2012) interpret investor sentiment as a broad and persistent tendency toward speculative behaviour, using Baker and Wurgler’s index. Schmeling (2009) agrees but uses survey-based sentiment indicators, such as consumer confidence indices, business sentiment surveys, and direct investor surveys, to reflect optimism or pessimism toward the market and economy. Tetlock (2007) extends sentiment measurement by analysing daily financial news coverage in the Wall Street Journal, using language tone to create a media sentiment index, linking information, environment, and media reporting to aggregate investor mood. Brown and Cliff (2005) adopt a qualitative approach in “Investor Sentiment and Asset Valuation,” using survey-based measures from the American Association of Individual Investors (AAII) and Investors Intelligence (II), where investors directly report bullish or bearish sentiments.

2.3.1 U.S. EVIDENCE

Despite different measurement methods, studies consistently indicate that investor sentiment affects stock returns. Baker and Wurgler’s (2006) foundational work, “Investor Sentiment and the Cross-Section of Returns,” finds that sentiment influences cross-sectional returns differently across firm characteristics in the U.S. market from 1963–2001. Small stocks show excess

returns during low sentiment, while the size effect reverses during high sentiment. Beyond size, younger firms outperform older ones, high-volatility firms outperform low-volatility ones, unprofitable firms outperform profitable ones, and non-dividend-paying stocks outperform dividend payers during low sentiment, with patterns reversing during high sentiment. It is important to note that many of these effects are found in samples extended back to 1935 and were supported by survey-based results in a 2007 review (Baker and Wurgler 2006; Baker and Wurgler 2007).

2.3.2 INTERNATIONAL EVIDENCE

Stambaugh et al. (2012) find that investor sentiment predicts anomalies in stock returns in the U.S. market from 1965–2008, with higher returns following low sentiment and lower returns following high sentiment, especially for hard-to-value and hard-to-arbitrage stocks. Schmeling (2009) extends this to 18 industrialized countries from 1973–2007, showing that sentiment significantly predicts stock returns internationally, with stronger effects in smaller, less developed markets and for hard-to-value stocks.

Baker et al. (2012) analyse data from 40 countries from 1994–2007, using country-specific sentiment indices based on Baker and Wurgler’s approach. Their panel regressions reveal that local and foreign sentiment affect national returns, with sentiment shocks in one country spilling over to economically and financially linked countries. Tetlock (2007) finds that increased media pessimism in the U.S. market from 1984–1999 leads to lower returns the following day, with partial reversals in subsequent days, confirming media sentiment as a driver of short-term market movements. Brown and Cliff (2005) link high sentiment to overvalued stock prices and lower following returns, noting that sentiment-driven mispricing continues due to limited arbitrage.

2.3.3 CONTRARIAN FINDINGS

However, some studies present contrary evidence. Brown and Cliff (2004) argue that sentiment’s effects on stock returns are modest and short-lived, with no significant predictive power beyond a few days to weeks after controlling for fundamentals. Garcia (2013) studies the U.S. market from 1990 to 2011 using a news-based sentiment measure, finding that sentiment’s predictive power is strong during economic expansions but weak or not existing

during recessions. This suggests that sentiment-driven mispricing may be outweighed by fundamentally driven investment choices during crises.

Beyond the U.S. market, several studies report varying or insignificant sentiment effects across different regions. González-Izquierdo et al. (2023) examine Eurozone countries from 2002 to 2020, finding that sentiment's impact varies by country and sector, with no statistically significant effect in many sectors after controlling for fundamentals, highlighting that sentiment is context dependent. Floros (2008) finds no causal relationship between survey-based sentiment and stock returns in the UK market from 1988 to 2004. Verma and Verma (2007) report similar findings for the U.S. market from 1977 to 2005, noting a weak or non-existent link between sentiment and market risk. This suggests that sentiment proxies may capture liquidity or institutional activity rather than investor mood. The existing literature shows that investor sentiment is an important but time-varying driver of cross-sectional stock returns, with effects that differ across periods, firm characteristics, and markets., its effect varies across time periods, firm characteristics and geographically.

3. DATA AND METHODOLOGY

The data and methodology section describes the data sources as well as the data preparation and construction procedures. It also presents the descriptive statistics of the sample and outlines the empirical methodology.

3.1 DATA SOURCES AND SAMPLE CONSTRUCTION

The data used in this analysis was obtained from various sources. The investor sentiment data based on Baker and Wurgler (2006) is retrieved from NYU Stern, where the sentiment index is publicly available.¹ Return data was downloaded from the Center for Research in Security Prices (CRSP) via Wharton Research Data Services (WRDS), and accounting data was obtained from Compustat, also via WRDS. Additionally, the returns for the Fama and French factors were collected from the Kenneth R. French Data Library.² The empirical analysis covers the

¹ Malcolm Baker and Jeffrey Wurgler, Investor Sentiment in the Stock Market, *Journal of Economic Perspectives* 21, no. 2 (2007): 129–152. Data available at NYU Stern. Accessed 22 November 2025. <https://pages.stern.nyu.edu/~jwurgler/>.

² Kenneth R. French Data Library, Fama/French Factors, Tuck School of Business, Dartmouth College. Accessed 22 November 2025. https://mba.tuck.dartmouth.edu/pages/faculty/ken.french/data_library.html.

period from January 2000 to December 2023. Sentiment data are available from December 1999 onward to allow for the construction of one-month lagged sentiment measures at the start of the sample.

3.1.1 SENTIMENT DATA

The sentiment measure used in this study follows the definition by Baker and Wurgler (2006) in their influential paper. Given the lack of standardized definitions and approaches to measure sentiment, as already discussed in the literature review, this specific approach was selected because the data is both recent and publicly available. The composite sentiment index constructed by Baker and Wurgler originally consisted of six different proxies: the closed-end fund discount, NYSE share turnover, the number and first-day returns of IPOs, the equity share in new issues, and the dividend premium. However, in the publicly available sentiment index, NYSE share turnover was excluded. This exclusion is justified by the rise of institutional high-frequency trading and the movement of trading activities away from traditional stock exchanges toward alternative platforms. Therefore, the final sentiment index consists of the five remaining indicators, which will be explained in more detail in the following paragraph.

The value-weighted dividend premium is included as a 12-month lagged ($t - 12$) value, following the methodology introduced by Baker and Wurgler (2004). In their paper, the dividend premium is defined as the logarithmic difference between the value-weighted average market-to-book ratios of dividend-paying and non-dividend-paying firms for each year. The monthly first-day returns on IPOs are calculated according to Ibbotson et al. (1994) with recent data updates sourced from Jay Ritter's website.³ Here, a volume-weighted average over the preceding 12 months is applied to smooth out short-term fluctuations, again using a lag of 12 months ($t - 12$). Similarly, the monthly IPO volume follows Ibbotson et al. (1994) and is likewise updated from Jay Ritter's dataset, employing a rolling sum of IPO volume over the previous 12 months.

The closed-end fund discount was sourced from Herzfeld for the period of 1987 to 2010, and subsequently from Morningstar starting in 2011.⁴ It is measured as the equal-weighted

³ Jay R. Ritter, IPO Data, Warrington College of Business, University of Florida. Accessed 22 November 2025. <https://site.warrington.ufl.edu/ritter/ipo-data/>.

⁴ Baker and Wurgler are using Herzfeld's Closed-End Fund Average for the period of 1987 – 2010. This average is a weekly publication in Barron's and provided by Thomas J. Herzfeld Advisors. The average discount or premium of U.S. closed-end equity funds relative to their net asset value is measured and reported as an index.

unlevered general equity discount, representing the average difference between the net asset values of closed-end stock fund shares and their market prices. The final component of the sentiment index, the equity share in new issues, is adopted from Baker and Wurgler (2000). It is defined as the ratio of equity issuance to the combined total of equity and debt issuance over the previous 12 months, with data drawn from the Federal Reserve Bulletin.⁵

The sentiment index data used in this thesis covers monthly observations from December 1999 to December 2023, the most recent available date at the time of this analysis. Subsequently, the data was standardized using the mean (μ) and the standard deviation (σ) for the full sample period, following the standardization approach introduced by Baker and Wurgler (2006). This resulted in the standardized sentiment index (z_{index}), defined as:

$$z_{index} = \frac{sentiment_{index_t} - \mu}{\sigma}$$

Here, μ denotes the average sentiment index over the entire sample period, and σ represents the corresponding standard deviation. Standardizing the index ensures consistent interpretation of sentiment changes over time and enables its use as a continuous variable in regressions. Additionally, the standardized index makes a straightforward classification of market phases, into "high" sentiment (positive z-scores) and "low" sentiment (negative z-scores), possible. In total, the dataset contains 289 observations with a maximum sentiment value of 2.73 and a

From 2011 onwards Baker and Wurgler are using Morningstar data on unlevered general equity closed-end funds.

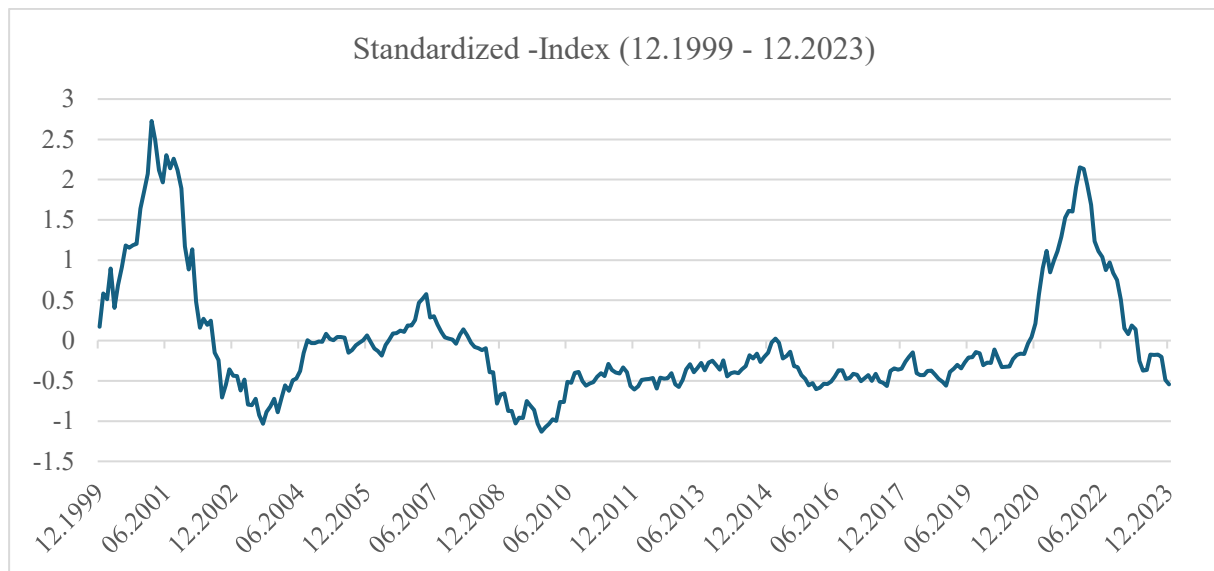
⁵ Equity share in new issues ($E/(E + D)$), based on aggregate "New Equity Issues" and "New Debt Issues" data from the Federal Reserve Bulletin. The transcribed series are available through Jeffrey Wurgler's data archive. Accessed 22 November 2025. <https://pages.stern.nyu.edu/~jwurgler/data/>

minimum value of -1.13. Sorting the dataset according to sentiment phases results in 91 observations for high sentiment ($z > 0$) and 198 observations for low sentiment ($z < 0$).

Figure 1

Standardized investor sentiment index (2000-2023)

This graph shows fluctuations in investor sentiment over time, with positive values (>0) indicating high-sentiment and negative values (<0) indicating low sentiment.



Source: Author's calculation based on data from Baker and Wurgler Sentiment Index.

When observing the magnitude of the sentiment index, it is notable that positive spikes are significantly larger than negative ones. This asymmetry is related to the construction of the sentiment index and the nature of several underlying components. Specifically, indicators like IPO first-day returns and IPO volumes can theoretically increase indefinitely, but they cannot drop below zero, making them positively skewed. A similar logic applies to the equity share in new issues, which has a natural lower bound of 0% and an upper bound of 100%, resulting in a positive skew as well.

Analysing the historical data, several visible spikes stand out from the average trend. The first significant spike occurred between 2000 and 2002, with a steep surge peaking in 2001 during the dotcom bubble. Following the bubble's burst, the index sharply declined, reaching approximately -1, reflecting investors' increased scepticism and pessimism after the market crash. Another noticeable positive spike appeared in 2007, shortly before the global financial crisis of 2008. This spike was succeeded by a substantial decline from 2008 through 2010, a period dominated by market pessimism, risk-aversion, and uncertainty among investors. A further notable increase occurred between 2020 and 2022, in the aftermath of the COVID-19

crisis. Following an initial sharp drop, a rapid recovery took place, driven primarily by monetary easing, fiscal stimulus, and renewed investor optimism.

3.1.2 RETURN DATA

The monthly return data used in this analysis covers the period from December 1999 to December 2023. The data was retrieved from CRSP and includes the following key variables: the unique firm identifier (PERMNO), monthly returns (RET), delisting returns (DLRET), stock price (PRC), shares outstanding (SHROUT), and exchange code (EXCHCD).

To clean and prepare the dataset, several filters were applied. Only stocks listed on the NYSE, AMEX, and NASDAQ were retained by selecting observations with exchange codes 1, 2, and 3. Additionally, only ordinary common shares were kept, corresponding to share codes 10 and 11. Observations with missing returns or market equity, as well as extreme returns below – 100%, were excluded to avoid incorrect.

In the process of the analysis further key metrics are calculated. This includes the market equity (ME):

$$ME = \frac{|PRC| * SHROUT}{1,000}$$

where the absolute value ensures that negative prices (due to data conventions) are corrected, and the result is expressed in millions of dollars. Adjusted returns (adj_{ret}) were calculated to account for potential delisting effects using the standard CRSP formula:

$$adj_{ret} = RET + (1 + RET) * DLRET.$$

This correction ensures that the total return reflects both regular performance and losses caused by delisting, which would otherwise lead to biased return estimates.

Volatility was proxied by 12-month rolling window standard deviation of log returns ($vol12$), following Baker and Wurgler (2006). This variable is later used for sorting firms by risk and controlling for it in analysis.

For momentum, the classical definition from financial literature was applied. The 12-month momentum was calculated as the cumulative return from month $t - 12$ to $t - 2$ (excluding $t - 1$ to avoid short-term reversals):

$$mom12 - 2_{i,t} = \sum_{r=t-12}^{t-2} r_{i,t}.$$

An additional momentum measure was included based on MSCI's methodology.⁶ It combines two periods t-12 to t-2 month and t-6 to t-2 month, which represents the momentum of the past 12 to the past 2 months and the past 6 to the past 2 months, and adjusts them for risk using their respective standard deviations:

$$mom12 - 2_{adj} = \frac{mom12-2}{sd12-2} \text{ and } mom6 - 2_{adj} = \frac{mom6-2}{sd6-2}$$

and then computes a final composite score as:

$$mom_{ishares} = \frac{1}{2} * (mom12 - 2_{adj} + mom6 - 2_{adj}).$$

Firm age was calculated using the first appearance of each PERMNO in the CRSP database (available from 1925 onward). From that date onward, the number of months since the first appearance of the respective company was tracked, assigning an age value as of December 2023 or the last valid observation. The measure therefore reflects a firm's listed age, the duration of its presence in the public markets, rather than its founding age.

3.1.3 ACCOUNTING DATA

The accounting data used in this analysis serves to characterize the firms in the sample based on fundamental financial metrics. Quarterly data was employed for the sample period from December 1999 to December 2023. While Baker and Wurgler (2006) rely on annual accounting figures, the use of quarterly data in this study is motivated by its greater timeliness and relevance. Since investors increasingly base decisions on quarterly earnings reports, this granularity offers a more realistic reflection of available information.

The following variables are included in the downloaded data: Total Assets (*atq*), common ordinary equity (*ceqq*), total liabilities (*ltq*), gross property, plant & equipment (*ppegqtq*), net property, plant & equipment (*ppent*), preferred stock (*pstk*), retained earnings (*re*), deferred taxes (*txdbq*), deferred taxes plus investment tax credit (*txdb*), common dividends (*dvc*), total dividends (*dvt*), net income (*ni*), net sales (*sale*), R&D expense (*xrd*), capital expenditure

⁶ See MSCI (2024). *MSCI Momentum Indexes Methodology*. The index methodology defines the momentum signal as a composite of risk-adjusted 6-month and 12-month price performance. Available at: <https://www.msci.com/index/methodology/latest/Momentum> (Accessed 22 November 2025).

(*capx*), common shares outstanding (*csho*) and employees (*emp*). With the use of these variables derived ratios are constructed from the raw items. These are:

- Book equity (BE): $BE = ceqq + txdbq + txditcq - pstkq$, which represents the accounting value of shareholders' claim.
- Book-to-market-ratio (BM): $BM = \frac{BE}{ME}$, to classify growth stocks versus value stocks.
- Return on assets (ROA): $ROA = \frac{niq}{atq}$, giving insights about the operating profitability,
- Capital-expenditure to asset ratio (cap_{at}): $cap_{at} = \frac{capxy}{atq}$, reflecting the investment intensity.

These ratios are used throughout the empirical analysis for portfolio sorts and regression specifications. The dataset was merged using PERMNO identifiers, resulting in a final panel comprising the following characteristics.

3.2 DESCRIPTIVE STATISTICS

Table 1

Summary statistics of the dataset after data cleaning

The table reports the number of observations, mean, standard deviation, minimum, percentiles (25th, 50th and 75th) and maximum values for the main firm level variables used in the analysis.

	count	mean	sd	min	p25	p50	p75	max
Market Equity	901486	5031.402	19439.59	1.52985	105.105	512.8138	2319.373	368582.8
Volatility	901486	.1422849	.0907985	.0245483	.0790165	.1174079	.17856	.6638776
Return on Assets	901486	-.0137228	.0796269	-1.16635	-.0139952	.0062747	.0185076	.198728
Adjusted Return	901486	.010159	.1881533	-.9936	-.0700935	.0028805	.075	17.60465
CapEx / Assets	901486	.028792	.0409474	.0000241	.0054841	.0148878	.0346763	.3888372
Listed Firm Age (years)	901486	26.99162	19.03264	1.5	12.5	23.75	35.08333	98.91666

Source: Author's calculation based on data from CRSP, Compustat, and the Baker and Wurgler Sentiment Index.

The dataset consists of approximately 900 thousand monthly firm-level observations, each including stock returns, sentiment indicators, and fundamental accounting variables, matched

via the PERMNO identifier. On average, firms in the sample have a market equity of USD 5 billion, with a strongly right-skewed distribution (maximum USD 368.6 billion) and a high standard deviation of USD 19.4 billion, indicating a high variation in firm size.

Return on assets (ROA) averages -1.4% , suggesting that, on average, companies recorded losses over the sample period. Adjusted returns average 1.0% , with a relatively high standard deviation of 18.8% . Volatility, measured as the 12-month rolling standard deviation of returns, averages 14.2% , indicating considerable variation in returns. The CapEx/Assets ratio stands at 2.9% , reflecting modest investment intensity. On average, firms are 27 years old, with ages ranging from just above one year to nearly 99 years.

A comparison of firm fundamentals during high and low sentiment periods reveals notable differences. Firms observed during low-sentiment phases tend to be older (average 28 years vs. 26 years) and larger (mean market equity: USD 5.04 billion vs. 5.01 billion) than those in high-sentiment phases. Companies also report higher profitability and return in low sentiment phases. In contrast, firms during high-sentiment periods show greater volatility (15.7% vs. 13.5%) and lower ROA (-1.9% vs. -1.1%), which may be attributed to increased investor risk appetite and the presence of weaker fundamentals in such market environments.

Table 2
Summary statistics sorted by sentiment_{t-1}

The table reports descriptive statistics for firm-level variables under periods of low and high investor sentiment. Variables include Market Equity, Volatility, Return on Assets, Adjusted Return, Capital Expenditure to Assets (CapEx/Assets), and Listed Firm Age. Presented statistics are the number of observations, mean, standard deviation, minimum, and maximum for each variable.

Low Sentiment	count	mean	sd	min	max
Market Equity	609471	5044.025	18588.59	1.52985	368582.8
Volatility	609471	.1350539	.0834497	.0245483	.6387326
Return on Assets	609471	-.0112558	.0726918	-.6588797	.1906543
Adjusted Return	609471	.0131944	.178983	-.9936	15.98446
CapEx/Assets	609471	.0284311	.0401063	.0000241	.3510943
Listed Firm Age (years)	609471	27.66262	18.98474	2.083333	98.91666

High Sentiment	count	mean	sd	min	max
Market Equity	292015	5005.057	21105.49	1.52985	368582.8
Volatility	292015	.157377	.1028596	.0245483	.6638776
Return on Assets	292015	-.0188715	.0922279	-1.16635	.198728
Adjusted Return	292015	.0038236	.205838	-.981295	17.60465
CapEx/Assets	292015	.0295452	.0426396	.0000298	.3888372
Listed Firm Age (years)	292015	25.59115	19.05628	1.5	98.91666

Source: Author's calculation based on data from CRSP, Compustat, and the Baker and Wurgler Sentiment Index.

These patterns support the hypothesis that investor sentiment is associated with differences in firm characteristics and market behaviour. The structure of the dataset, comprising a broad cross-section and a long time series (2001–2023), provides a solid foundation for robust testing of sentiment-driven return effects.

Due to the strong right skew in the distributions of Market Equity and Volatility, key variables such as Size, Book-to-Market, Age, and CapEx/Assets were log-transformed to improve distributional symmetry. Additionally, yearly winsorization at the 0.5% and 99.5% percentiles was applied to reduce the influence of extreme outliers while keeping meaningful cross-sectional variation for Market Equity, Book-to-Market ratio, Return on Assets, CapEx/Asset ratio and Volatility. These data cleaning steps were essential to enhance the stability and interpretability of regression results.

3.3 EMPIRICAL METHODOLOGY

This section describes the steps used to test whether investor sentiment predicts return differences across firm characteristics. The empirical analysis proceeds in several stages. Portfolios are constructed and spreads formed, according to the process in the base paper of Baker and Wurgler (2006) with the following estimation of two complementary regression models.

3.3.1 SAMPLE PREPARATION AND PORTFOLIO CONSTRUCTION

The first step, after retrieving the data and calculating the relevant metrics and ratios is to merge the datasets. This is done via a WRDS merge using the identifiers PERMNO and GVKEY.

Sentiment data and Fama-French factor returns (monthly frequency) are then merged into the file. To focus the analysis on industrial firms, financial institutions and REITs are excluded by removing SIC codes 6000–6999 and 6798, respectively.

As explained in the accounting data part, book equity is used for several key ratios, therefore observations with missing or negative book equity are removed. Additionally, only firms with at least nine return observations in the prior 12 months are used, to ensure sufficient data for computing lagged variables.

The sentiment data is converted into a binary dummy variable, where:

- 1 = high sentiment (standardized sentiment index_{t-1} > 0).
- 0 = low or neutral sentiment (standardized sentiment index_{t-1} < 0).

Because accounting data are reported at a quarterly frequency while stock returns are monthly available, the analysis restricts returns to the first month following each firm's quarterly earnings announcement, identified by using CRSP announcement dates. This ensures that monthly returns are matched with the most recently available firm characteristics. As a consequence of this restriction, the effective sample for portfolio-level regressions comprises 94 monthly observations, corresponding to approximately one-third of the full 289-month sentiment time series.

The described steps enable the classification of the accounting data and returns data into a respective dummy of high or low sentiment. The next step consists of the portfolio construction and formation of long-short spreads.

3.3.2 TIME-SERIES SPREAD REGRESSION

To examine how investor sentiment affects return premia, firms are sorted each month into ten deciles based on eight characteristics.

- *Spread Size*: small versus large firms,
- *Spread B/M*: value versus growth firms,
- *Spread Profitability*: high versus low return on assets,
- *Spread Age*: young versus mature firms,
- *Spread CapEx/Assets*: firms with high versus low investment intensity,

- *Spread Volatility*: firms with high versus low past return volatility,
- *Spread Momentum (12 – 2)*: past winners versus losers,
- *Spread Momentum (MSCI-based)*: momentum signal based on MSCI index weights.

Average monthly returns are computed for each decile across the sample period and for each characteristic, the long–short spread is defined as the average return of the top three deciles (8–10) minus the average return of the bottom three deciles (1–3). This procedure produces eight characteristic-based long-short time series. Due to the quarterly frequency of accounting data and the restriction to using only the first month after each quarterly earnings announcement, the effective time-series sample for portfolio sorts and spread regressions consists of 94 monthly observations. These spreads form the dependent variables in the time-series regression models described below.

To test the significance of these spreads, time-series spread regressions are used following the approach of Baker and Wurgler (2006):

$$\text{univariate regression: } Spread_t = \alpha + \beta S_{t-1} + \varepsilon_t$$

$$\text{multivariate regression: } Spread_t = \alpha + \beta S_{t-1} + \sum F_{t-1} + \varepsilon_t$$

The unit of observation is one return spread per calendar month. The main explanatory variable is lagged sentiment from the previous month, applied as a dummy variable. Monthly Fama-French five factors (Rm-Rf, SMB, HML, RMW, CMA) and momentum (UMD) are used as controls. Factors such as SMB or HML are excluded when they overlap with the dependent spread, to avoid multicollinearity.

The coefficient β tests whether high investor sentiment in the previous period predicts systematic mispricing. For example, a positive β in the B/M spread indicates that value outperforms growth to a lesser degree in optimistic markets, suggesting increased mispricing of growth stocks. Given the limited time series (94 observations), statistical inference is based on bootstrapped standard errors with 200 replications.

3.3.3 FIRM-LEVEL PANEL REGRESSION

In addition to the time-series spread regressions a second regression approach is used. The firm-level panel regression examines whether sentiment interacts with firm characteristics in predicting stock returns, following phases of high or low sentiment. The dependent variable is

the firm-specific adjusted return. Independent variables include lagged firm characteristics (e.g., B/M, ME, ROA, volatility, momentum, and age). Sentiment is again lagged by one month and interacted with each firm characteristic.

Firm-level panel regression:

$$R_{it} = \alpha + \beta_1 S_{t-1} + \beta_2 x_{it-1} + \beta_3 S_{t-1} x_{it-1} + \varepsilon_t$$

Where S_{t-1} is the sentiment dummy, x_{it-1} is the firm characteristic, and β_3 captures the change in the return premium under high sentiment. β_2 reflects the average premium in neutral sentiment periods.

These two types of regressions complement one another. The time-series spread regression captures a macro-level view of aggregate mispricing across portfolios, while the firm-level panel regression offers a micro-level perspective by testing how sentiment affects return premia at the individual firm level, controlling for unobserved firm and industry heterogeneity.

4. EMPIRICAL RESULTS

The empirical results section covers the results of the two regression approaches used in this analysis. First, the portfolio-level results are presented, followed by the firm-level panel results.

4.1 DESCRIPTIVE PORTFOLIO EVIDENCE (UNCONDITIONAL)

After sorting the firms into deciles, long-short spreads are computed by taking the equal weighted returns of the top decile minus the bottom decile. The difference between the computed spreads in high- and low-sentiment phases are then compared.

Table 3

Summary of characteristic spreads by sentiment regime

The table presents the average characteristic spreads for low- and high-sentiment periods as well as the difference between the two ($spread_{diff}$). Variables include firm-level measures such as Size, Book-to-Market ratio,

Profitability, Firm Age, Capital Expenditure to Assets (CapEx/Assets), Volatility, and two momentum measures (Momentum and Momentum_{MSCI}). Detailed decile return tables are provided for illustration in Appendix II.

variable	spread _{low}	spread _{high}	spread _{diff}
Size (ln)	-0.005	0.018	0.023
Book-to-Market (ln)	-0.015	-0.032	-0.017
Profitability	0.003	0.015	0.012
Listed Firm Age (ln)	-0.008	0.008	0.015
CapEx/Assets (ln)	0.000	-0.003	-0.003
Volatility (ln)	0.002	-0.025	-0.027
Momentum	-0.007	0.008	0.015
Momentum _{MSCI}	-0.006	0.015	0.020

Source: Author's calculation based on data from CRSP, Compustat, and the Baker and Wurgler Sentiment Index.

The decile-based portfolio sorts reveal notable differences in return spreads between periods of high and low investor sentiment. Using Size(ln) as an example, the spread between big and small companies (small-firm portfolio minus large-firm portfolio) in low sentiment phases for monthly returns is -0.5% and in high sentiment phases +1.8%. This result implies that smaller firms slightly underperform in low-sentiment periods but outperform in high-sentiment periods. The spread difference of +2.3% indicates size is a firm characteristic which is strongly influenced by sentiment. The largest spread differences appear in volatility and size. While the size spread increases sharply in high-sentiment months, the volatility spread turns negative with a total difference of -2.7 percentage points from +0.2% in low-sentiment months to -2.5% in high-sentiment months, indicating that high volatile stocks perform worse when sentiment peaks.

The book-to-market ratio, which distinguishes value from growth stocks, consistently shows negative spreads. In low-sentiment phases, the spread is -1.5%, compared to -3.2% in high-sentiment phases. This implies a stronger preference for growth stocks when sentiment is optimistic, aligning with findings in the behavioural finance literature.

For profitability and both momentum measures, the spread differences are moderate but still existing. These suggest that more profitable firms with higher momentum tend to benefit during optimistic periods. Whereas in contrast, CapEx/Assets is not showing a notable spread, indicating that firms' investment intensity is largely unaffected by investor sentiment.

4.2 PREDICTIVE SPREAD REGRESSION (PORTFOLIO LEVEL)

To assess whether these spread differences are statistically significant, time-series regressions of spreads on lagged sentiment were conducted as described in the methodology section. The results confirm that the standardized sentiment index has a statistically significant impact on several of the characteristic-based spreads, supporting the hypothesis that sentiment contributes partly to the cross-section of returns.

Table 4

Univariate Regression of monthly characteristic spreads

*This table reports univariate regression of monthly characteristics spreads on lagged investor sentiment. Each column represents a separate regression for one firm characteristic (Size, Book-to-Market, Profitability, Age, CapEx/Assets, Volatility, Momentum, and Momentum_{MSCI}). Standard errors are reported in parentheses and *, ** and *** are showing significance at the 10%, 5% and 1% level. The standard errors are used to calculate the significance and the constant/intercept (_cons) predicts the value of the spread when sentiment is low (dummy variable is 0).*

	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
Spreads	Size	BM	ROA	Listed Age	CapEx/Assets	Volatility	Mom	Mom _{msci}
Sentiment	0.0192**	-0.0143***	0.0137**	0.0180***	-0.0019	-0.0335***	0.0135**	0.0163**
(SE)	(0.0086)	(0.0053)	(0.0068)	(0.0048)	(0.0064)	(0.0086)	(0.0066)	(0.0066)
_cons	0.0036	-0.0212***	0.0073	-0.0018	-0.0012	-0.0082	-0.0012	0.0019
(SE)	(0.0051)	(0.0046)	(0.0046)	(0.0040)	(0.0036)	(0.0085)	(0.0048)	(0.0046)
Obs.	94	94	94	94	94	94	93	93
R-squared	0.082	0.061	0.053	0.118	0.001	0.114	0.047	0.068

Source: Author's calculation based on data from CRSP, Compustat, and the Baker and Wurgler Sentiment Index.

Table 5

Multivariate Regression of monthly characteristic spreads

*This table shows the extension of the univariate regression, by controlling for the Fama-Five five-factors (Rm–Rf, SMB, HML, RMW, CMA) as well as the momentum factor (UMD). The coefficient on sentiment now captures how sentiment affects the spread after controlling for common risk factors. Empty spacing in the regression rows results from factor exclusions. To avoid overlap, the SMB factor is omitted in the size spread regression, HML in the book-to-market regression, and UMD in momentum regressions. Standard errors are clustered at the firm level. *, ** and *** indicate significance at the 10%, 5% and 1% levels.*

Spread	(1) Size	(2) BM	(3) ROA	(4) Listed Age	(5) CapEx/Assets	(6) Volatility	(7) Mom	(8) Mom _{msci}
Sentiment	0.0174** (0.0086)	-0.0110 (0.0071)	0.0066 (0.0070)	0.0091 (0.0057)	-0.0007 (0.0078)	-0.0130 (0.0089)	0.0089 (0.0084)	0.0087 (0.0080)
rmrf	-0.0030* (0.0016)	0.0019 (0.0013)	-0.0026** (0.0013)	-0.0017 (0.0011)	-0.0025* (0.0013)	0.0099*** (0.0015)	-0.0031* (0.0017)	-0.0048*** (0.0014)
hml	-0.0053** (0.0025)		0.0009 (0.0019)	0.0034* (0.0018)	0.0013 (0.0019)	0.0024 (0.0030)	-0.0020 (0.0024)	-0.0031 (0.0021)
rmw	-0.0017 (0.0027)	0.0040 (0.0028)	0.0023 (0.0028)	0.0021 (0.0021)	0.0041** (0.0021)	-0.0011 (0.0030)	0.0019 (0.0034)	0.0023 (0.0031)
cma	0.0048 (0.0037)	0.0026 (0.0024)	-0.0050* (0.0027)	-0.0002 (0.0025)	-0.0048 (0.0031)	-0.0025 (0.0045)	-0.0007 (0.0042)	0.0003 (0.0040)
umd	-0.0013 (0.0019)	-0.0028* (0.0016)	0.0026 (0.0017)	0.0010 (0.0012)	-0.0032* (0.0017)	-0.0009 (0.0019)		
smb		0.0034* (0.0019)	-0.0018 (0.0018)	-0.0025 (0.0016)	0.0013 (0.0022)	0.0060*** (0.0022)	0.0001 (0.0023)	0.0008 (0.0024)
_cons	0.0050 (0.0048)	-0.0236*** (0.0051)	0.0058 (0.0046)	-0.0034 (0.0041)	-0.0001 (0.0046)	-0.0090* (0.0051)	-0.0021 (0.0051)	0.0000 (0.0051)
Obs.	94	94	94	94	94	94	93	93
R-squared	0.234	0.274	0.317	0.278	0.154	0.659	0.189	0.364

Source: Author's calculation based on data from CRSP, Compustat, and the Baker and Wurgler Sentiment Index.

Using size as an example, the first row reports the estimated effect of sentiment on the return spread. A one-unit increase in sentiment is associated with a 0.0192 (~1.92%) increase in the

size-spread in the following month, meaning small firms are expected to outperform large firms after high-sentiment months.

Analysing the results across the spreads, a central takeaway is, that only the size-spread keeps its positive coefficient and remains statistically significant at the 5% significance level using the univariate and multivariate regression approach. This confirms the thesis of smaller firms outperforming larger firms in periods following high-sentiment, consistent with the literature reporting that investors are favouring smaller companies in high-sentiment phases.

In contrast, book-to-market represents a negative value, which is significant, but only in the univariate regression and loses the significance in the multivariate approach. Initially it seems that investors are favouring growth firms over value firms in high sentiment phases, but when factor controls are added this effect vanishes as it overlaps with existing risk-factor dynamics.

The spreads age and profitability are positive and significant in the univariate approach but are losing the significance with the inclusion of the risk-factors. These findings indicate that without controls, there is an investors' preference for profitable and older firms in high-sentiment phases. However, this effect becomes statistically insignificant with the inclusion of risk-factors.

The volatility is significantly negatively affected by sentiment. A higher sentiment index in the previous month predicts weaker performance for high-volatility stocks compared to more stable ones. This result indicates a negative relationship between sentiment and volatility-based return spreads. Instead, the findings suggest that investors in high-sentiment periods favour firms with lower risk profiles. One possible explanation is that high investor sentiment may be a delayed response to the strong past performance of less volatile firms, rather than an indication of greater willingness to invest in riskier companies. This effect also loses its significance in the multivariate approach.

No statistically significant link is observed between investor sentiment and the spreads based on CapEx/Assets and the two momentum measures. This indicates that, in this sample, the investment intensity and recent past performance are not systematically influenced by sentiment once risk factors are considered.

In summary, the univariate regressions indicate that lagged investor sentiment is associated with several characteristic-based return spreads. After adjusting for common risk factors, firm size remains the most robust sentiment-sensitive characteristic and reports statistical significance at

10%, 5% and 1% levels. Other characteristics, such as book-to-market and firm age, exhibit attenuated effects that fall short of standard significance thresholds in the risk-adjusted specifications.

4.3 FIRM-LEVEL EVIDENCE (PANEL REGRESSIONS)

To extend the analysis of return spreads at the portfolio level, firm-level panel regressions were conducted. While the decile-based analysis offers valuable insights into the average return differences across characteristics under different sentiment conditions, the panel regression allows for more precise control of unobserved firm-specific and industry-specific heterogeneity. By including firm and industry fixed effects, this approach enables a more accurate identification of systematic relationships between past sentiment and returns, as well as how sentiment interacts with individual firm characteristics.

Table 6**Firm-level panel regression with one month lagged sentiment**

*This table reports firm-level regression with one month lagged investor sentiment and lagged firm characteristics, including interaction terms between sentiment and each characteristic. Column (1) shows the regression without fixed effects, column (2) with firm fixed effects and column (3) with firm and industry fixed effects. Standard errors are clustered at the firm level. *, ** and *** indicate significance at the 10%, 5% and 1% levels.*

	(1)	(2)	(3)
	adj_ret	adj_ret	adj_ret
Sentiment _{t-1}	-0.0745*** (0.0219)	-0.0255 (0.0234)	-0.0263 (0.0236)
BM _{t-1}	0.0066*** (0.0025)	0.0117*** (0.0043)	0.0108** (0.0042)
Size _{t-1}	-0.0010 (0.0009)	-0.0110*** (0.0035)	-0.0112*** (0.0035)
ROA _{t-1}	0.3028*** (0.0778)	0.3047*** (0.0866)	0.3090*** (0.0868)
Volatility _{t-1}	0.0017 (0.0045)	0.0016 (0.0052)	0.0016 (0.0052)
Momentum _{t-1}	-0.0017 (0.0047)	0.0014 (0.0051)	0.0013 (0.0052)
Listed Age _{t-1}	-0.0023 (0.0030)	-4.7835 (3.2605)	-8.3895*** (2.2269)
Sentiment x BM	-0.0121*** (0.0046)	-0.0132*** (0.0043)	-0.0129*** (0.0043)
Sentiment x Size	-0.0040** (0.0020)	-0.0039* (0.0021)	-0.0039* (0.0021)
Sentiment x ROA	-0.2618*** (0.1004)	-0.2817*** (0.1038)	-0.2821*** (0.1040)
Sentiment x Volatility	-0.0257*** (0.0081)	-0.0204** (0.0084)	-0.0202** (0.0086)
Sentiment x Momentum	-0.0040 (0.0085)	-0.0077 (0.0091)	-0.0074 (0.0092)
Sentiment x Listed Age	0.0090* (0.0053)	-0.0006 (0.0060)	-0.0002 (0.0061)
Obs.	9110	9056	9055
Adj. R ²	0.009	0.033	0.033
Firm FE	No	Yes	Yes
Industry FE	No	No	Yes

Source: Author's calculation based on data from CRSP, Compustat, and the Baker and Wurgler Sentiment Index.

The table presents the results of three panel regression specifications, aiming to examine the effect of monthly lagged sentiment (Sentiment_{t-1}) and firm characteristics on firms' subsequent monthly stock returns. The three models differ by the inclusion of fixed effects: model one includes no fixed effects, model two includes firm fixed effects, and model three includes both firm and industry fixed effects. In model (1), higher sentiment levels in the previous months are associated with significantly lower returns in the following month. Once fixed effects are added in model (2) and (3), this effect loses its significance. Without fixed effects, the significant coefficient amounts to -0.0745, while its impact decreases and becomes insignificant once firm and industry fixed effects are included.

The coefficients on the lagged firm characteristics also reveal meaningful patterns. A positive and significant coefficient in all models for the log of the book to market ratio indicates that value stocks tend to outperform growth stocks in this sample. In contrast, size has a negative effect and becomes significant once fixed effects are included, which is consistent with the small firm premium, meaning that smaller companies yield higher average returns than larger ones. Profitability shows a strong positive relationship with returns, significant through all models, suggesting that more profitable firms earn higher returns. Volatility has a small and not significant impact, implying no correlation between past volatility and returns. Momentum shows no significant predictive power for returns, and firm age is only significant in the third model, however the unusual large coefficient, suggests that this is a scaling or multicollinearity rather than an economically meaningful effect.

The interaction terms provide insight into how sentiment changes the return characteristic relationship. Significant negative interaction effects are observed between sentiment and book to market, size, profitability and volatility. This indicates that in high-sentiment periods, growth and small firms perform better, while profitable and high volatility stocks underperform. Because Size is measured as log market equity, the negative $\text{Sentiment} \times \text{Size}$ coefficient implies that returns decline more with sentiment as firm size increases, i.e., smaller firms benefit disproportionately in high-sentiment periods. The interaction with momentum is not significant, suggesting that past performance does not meaningfully influence the relation of returns to sentiment. A weak interaction between firm age and sentiment is observable across all models, implying no clear effect of sentiment on age-related performance differences.

Overall, the panel regression confirms that sentiment significantly affects stock returns and interacts with several firm characteristics. The findings highlight that the relationship between

firm characteristics and returns depends on sentiment conditions, but the strength of this relationship varies across characteristics and becomes clearer when firm and industry level heterogeneity is considered. Small and growth-oriented firms benefit more from high-sentiment periods, while highly profitable and volatile firms tend to underperform.

4.4 SYNTHESIS OF PORTFOLIO-LEVEL AND FIRM-LEVEL EVIDENCE

When comparing the portfolio-level spread regressions with the firm-level panel regressions, a consistent finding is that investor sentiment primarily affects the size dimension, with smaller firms outperforming larger firms following high-sentiment periods in both approaches. Beyond size, the results differ across methodologies. Portfolio-level regressions show temporary relationships between sentiment and return spreads based on age, profitability, and volatility, but these effects largely disappear once common risk factors are controlled for. Firm-level panel regressions provide a more detailed perspective, showing that high-sentiment periods are associated with stronger performance of smaller and more growth-oriented firms, while highly profitable and high-volatility firms tend to underperform. Taken together, both approaches suggest that sentiment-driven mispricing is most visible among small and growth-oriented firms. The discrepancy between portfolio-level and firm-level book-to-market results can be summarized as follows. Firm-level regressions show that value stocks earn higher returns on average, as reflected in the positive book-to-market coefficient. However, the negative interaction between sentiment and book-to-market suggests that this value premium weakens during high-sentiment periods. At the portfolio level, this appears as the relative outperformance of growth stocks during optimistic market phases.

5. DISCUSSION

This chapter discusses the empirical findings in relation to existing literature and the study's research questions. It interprets the main results, compares them with the findings of Baker and Wurgler (2006), and places them in the context of changes in financial markets. The chapter further addresses practical implications, limitations of the analysis, and suggestions for future research.

5.1 INTERPRETATION OF MAIN FINDINGS

The empirical results of this work provide new insights into the role of investor sentiment in explaining the cross-section of stock returns. They partly support core findings from Baker and Wurgler (2006) which serve as the foundation of this study but also reveal deviations from classical behavioural finance theory. These deviations can be attributed to methodological differences as well as structural changes in financial markets over the past two decades.

At the firm level, a significant negative relationship between the lagged sentiment variable and the monthly stock returns appears only in the specification without fixed effects. The finding partially and in a weak form confirms the hypothesis that high investor sentiment leads to overpricing in the short term, followed by return corrections. This observation is consistent with prior literature. Barberis and Thaler (2003) and Stambaugh et al. (2012) also show that periods of high sentiment are followed by lower returns, especially for stocks that are difficult to value or arbitrage. However, the sentiment return relation is not robust across all specifications, the effect disappears once firm and industry fixed effects are applied. This suggests that return predictability may be partially driven by unobserved firm-specific heterogeneity rather than pure sentiment-driven mispricing.

At the portfolio level, sentiment shows a meaningful influence on characteristic-based spreads. The most notable finding is that small firms outperform large firms more strongly in high-sentiment periods, mostly consistent with the existing literature. This is broadly consistent with behavioural theories, that state sentiment-driven mispricing is more present in companies that are hard to value. The remaining characteristics show considerably weaker correlation with sentiment. Book-to-market and volatility are showing signs of sensitivity to sentiment, but these effects disappear once standard risk factors are included, indicating, they are not independently sentiment driven.

The firm-level regressions offer more detailed insights into how sentiment changes the relationship between firm characteristics and returns. Consistent with the findings of Baker and Wurgler (2006), firms with lower book-to-market ratio (i.e., growth stocks) show higher positive performance in high-sentiment phases. In contrast to the spread regressions, the effect remains significant after controlling for fixed effects, indicating that sentiment primarily influences the value–growth dimension at the firm level rather than through portfolio spreads. These findings suggest that investor optimism continues to favour growth-oriented firms, while value-oriented companies are less rewarded. One possible explanation is a shift in investor

preferences. In a post-crisis environment characterized by uncertainty, investors may still seek growth opportunities during optimistic periods, but in a more risk-aware way. In addition, the interactions for momentum and firm age are not consistently positive, indicating that sentiment-driven investors are not necessarily drawn to established firms with long track records, but rather to firms that signal growth potential and align with prevailing market optimism. By contrast, volatility interacts negatively with sentiment in the firm-level panel regressions. However, this effect does not remain robust at the portfolio level once common risk factors are taken into account. High-volatility firms tend to underperform during high-sentiment periods, but the lack of consistent portfolio-level evidence calls for caution when interpreting this result as a general risk-seeking channel. Instead, it suggests that periods of strong sentiment may represent the peak of a market cycle, where riskier assets begin to underperform as sentiment anticipates future corrections. Finally, the reduced significance of profitability and firm size after including fixed effects indicates that unobserved firm specific characteristics, such as strategic positioning, innovation capacity, or market dominance, may absorb part of the explanatory power of these variables.

5.2 COMPARISON WITH BAKER AND WURGLER (2006)

Several factors could explain these differences. First, the analysed periods differ. This thesis covers the years 2000 to 2023, which were marked by multiple severe crises and global shocks. These include the dotcom bubble, the global financial crisis, and the COVID-19 pandemic. Each crisis had its own characteristics, with changing investor reactions and evolving patterns in recovery dynamics. Second, the rise of institutional trading, algorithmic decision making, and tighter regulation may have weakened the behavioural biases documented in earlier decades. Third, differences in methodology may also contribute to the variation in results, particularly the use of quarterly accounting data, which provides more timely information than annual data and may better capture investor behaviour in real time. These factors help explain why sentiment effects observed in earlier decades differ from those in the post 2000 period. In particular, the effects related to firm age, profitability, and volatility are less pronounced than in Baker and Wurgler's (2006) analysis.

5.3 SENTIMENT IN THE FACE OF MARKET EVOLUTION AND CRISES

The crises themselves illustrate how market behaviour and sentiment have changed. The dotcom crash between 2000 and 2002 was driven largely by retail investor speculation in internet stocks and led to an extended bear market, with the S&P 500 taking nearly seven years to recover. The 2008 crisis was system-wide and originated in credit markets, leading to a broad sell-off, with full recovery taking about four years. In contrast, the COVID-19 crisis of 2020 was an exogenous shock. It caused a sharp but short-lived downturn, and the S&P 500 returned to its pre-crisis level within five months. This rapid rebound was largely driven by coordinated fiscal and monetary policy. The sentiment index reflects this development, as sentiment dropped sharply during COVID-19, it rebounded just as quickly, aligning with the V-shaped recovery in stock prices and macroeconomic expectations.

5.4 PRACTICAL IMPLICATIONS

From a practical perspective, these findings suggest that sentiment remains a relevant signal for investment strategies. Even though the outcomes differ from classical predictions, sentiment remains a driver of relative performance. Investors and asset managers can use sentiment indicators in portfolio construction, particularly by overweighting firms that tend to benefit from high-sentiment periods in this sample, most notably smaller and growth-oriented firms. By contrast, caution may be warranted when allocating capital to highly volatile firms during optimistic phases, as these firms may be more prone to corrections. However, volatility loses significance in multivariate spread regressions, indicating that while sentiment affects the volatility–return relation at the firm level, volatility does not exhibit a robust standalone response to sentiment once common risk factors are controlled for.

5.5 LIMITATIONS AND FUTURE RESEARCH

5.5.1 SAMPLE AND DATA LIMITATIONS

The obtained results and their interpretation are subject to several limitations. Due to data availability the analysis is focussed on the US-Market, extending the findings to other global markets would require additional investigation. Moreover, the sample period of this study differs substantially from Baker and Wurgler's (2006) sample period. The recent period is characterised by major structural changes and several severe crises, including the aftermath of the dot-com bubble, the global financial crisis, low-interest rate environments and the COVID-

19 shock. In addition, quantitative and algorithmic trading system have altered market dynamics. Since such trading strategies are not sentiment driven, they can dilute the sentiment driven reactions. Also, from the methodological perspective limitations must be noted. Likewise, Baker and Wurgler equal-weighted portfolios are used for decile-sorts and spreads, which allocates weight towards small firms over proportionally in the portfolio. This can increase the sentiment effect for the sample, because smaller firms react are more sensitive to sentiment. From a data perspective, an important limitation concerns the measurement of firm age. The variable age captures the number of years since a firm's stock market listing rather than its economic age. Accordingly, all age-related findings should be interpreted in this context. A critical limitation of the firm-level panel regressions is the substantial reduction in sample size from 901,486 firm-month observations to 9,110 observations. This reduction results from the restriction to post-earnings months, missing accounting data, and the requirements of fixed-effects estimation. As a consequence, the panel sample may overrepresent larger and more established firms with continuous data coverage, potentially attenuating sentiment effects if smaller and more sentiment-sensitive firms are disproportionately excluded.

5.5.2 METHODOLOGICAL CONSTRAINTS

Furthermore, the study may be subject to reverse causality, as stock market performance itself can influence investor sentiment. For example, firms are more likely to perform initial public offerings when market valuations are high. Since both IPO volume and first-day IPO returns are components of the sentiment index, increased sentiment may partly reflect favourable market conditions rather than act as an independent predictor. In addition, fixed effects remove time-invariant firm characteristics. For variables such as firm size, the limited variation within firms over time reduces cross-sectional variation and weakens the estimated effects.

5.6 DIRECTIONS FOR FUTURE RESEARCH

When conducting further research in this field, it would be useful to expand into different markets and asset classes. The results of this study could be tested globally; although existing studies cover international financial markets, many of them rely on older samples and would benefit from being updated. Within the last decade the equity markets changed by a lot and new assets emerged for retail investors to allocate their money in. Especially cryptocurrencies gained a lot in popularity and became more important in retail portfolios. Beyond global and

asset-class expansion, varying approaches for the construction of the sentiment index should be considered, for example the number of creations of new crypto coins as well as return data of cryptocurrencies should be included. With the use of brokerage platforms asset allocation of retail investors can be analysed as well and give potential insights of investor sentiment. Further factors should be considered as well testing the dependence of firms on investor sentiment. In this work quantitative factors have been used to explain the cross-section of returns. Here more qualitative factors should be included, like news coverage or social media sentiment. This can explain possibly some more cross-sectional returns. Methodologically, because sentiment is asymmetrical and nonlinear, threshold or quantile-based regressions can capture sentiment effects more precise.

6. CONCLUSION

This dissertation analyses the relation between investor sentiment and the cross-section of returns for the post-2000 period. Building on Baker and Wurgler's (2006) returns in the U.S. equity market are analysed. The selected period is characterized by several market anomalies, like global crises and geopolitical tensions, which motivates the reassessment of the existing literature.

The empirical approach combines two empirical methodologies. The first one is used according to Baker and Wurgler's (2006) methodology. Firms are sorted into decile portfolios based on their characteristics, and a time-series regression of characteristic return spreads on lagged investor sentiment are performed. The second approach estimates how sentiment interacts with firm characteristics while controlling for unobserved firm and industry heterogeneity. Investor sentiment is measured using the Baker and Wurgler sentiment index and all regressions control for common risk factors, including the Fama-French five factors and momentum.

The results show that investor sentiment has a negative relationship with future stock returns in simple specifications. Higher sentiment levels predict lower future returns, suggesting that optimistic market conditions are associated with overvaluation that is corrected over time. However, this effect disappears once firm and industry fixed effects are included. This indicates that sentiment's predictive power mainly reflects cross-sectional differences between firms rather than time-series variation within firms. Accordingly, cross-sectional return premia differ across sentiment regimes. Decile-based portfolio sorts reveal clear differences in return spreads between high- and low-sentiment periods, indicating that sentiment influences how investors

price certain firm characteristics. However, both the strength and robustness of these effects vary considerably across characteristics. Overall, the hypothesis that speculative firms earn higher returns during high-sentiment periods receives only limited support. While univariate regressions show sentiment-related return patterns for characteristics typically associated with speculative firms, such as low profitability, high volatility, and young firm age, these effects largely disappear once standard risk factors are included. This suggests that they overlap with established risk-based explanations rather than representing independent sentiment effects. While firm size remains the most robust characteristic for sentiment-driven mispricing across both portfolio- and firm-level analyses, firm-level regressions also indicate persistent sentiment interactions with the book-to-market ratio. This suggests that value–growth dynamics continue to be affected by investor sentiment at the firm level, even when portfolio-level effects weaken. This result confirms that at least part of the classical evidence from Baker and Wurgler (2006) still holds in modern markets.

These findings contribute to the literature in several ways. They provide updated evidence on sentiment effects using a recent sample period. By combining portfolio level spread regressions with firm level panel regressions, the study offers a diverse perspective on sentiment effects. The results also clarify which sentiment related patterns remain economically and statistically relevant today. However, sentiment effects appear more selective and less substantial than earlier studies suggested. The weakening of many classical sentiment effects may reflect increased market efficiency, improved information availability, and greater arbitrage activity by institutional investors, which likely constrain sentiment driven mispricing, particularly for firm characteristics linked to risk factors.

Several limitations should be noted. The analysis focuses exclusively on U.S. equity markets and relies on an aggregate sentiment measure. Linear model specifications may not fully capture nonlinear or asymmetric sentiment effects, especially during extreme periods. Even though the sentiment measure is lagged, concerns about endogeneity and reverse causality cannot be completely excluded.

Future research could analyse sentiment effects in international markets and across alternative asset classes like cryptocurrencies. Incorporating qualitative sentiment measures from news coverage or social media data could improve understanding of sentiment driven behaviour. More flexible models that allow sentiment effects to differ across market conditions (nonlinear) may better capture sentiment dynamics in modern financial markets.

In conclusion, investor sentiment continues to influence the cross-section of stock returns, but its impact is more nuanced than previously documented. While firm size remains a robust sentiment sensitive characteristic in the 2000–2023 period, other characteristics show weaker effects relative to those documented by Baker and Wurgler (2006) for 1963–2001. Whether this reflects genuine market evolution or differences in sample construction and methodology remains an open question for future research.

REFERENCES

- Baker, Malcolm, and Jeffrey Wurgler (2000). “The Equity Share in New Issues and Aggregate Stock Returns.” *Journal of Finance* 55 (5): 2219–2257.
- Baker, Malcolm, and Jeffrey Wurgler (2004). “Appearing and Disappearing Dividends: The Link to Catering Incentives.” *Journal of Financial Economics* 73 (2): 271–288.
- Baker, Malcolm, and Jeffrey Wurgler (2006). “Investor Sentiment and the Cross-Section of Stock Returns.” *Journal of Finance* 61 (4): 1645–1680.
- Baker, Malcolm, and Jeffrey Wurgler (2007). “Investor Sentiment in the Stock Market.” *Journal of Economic Perspectives* 21 (2): 129–152.
- Baker, Malcolm, Jeffrey Wurgler, and Yu Yuan (2012). “Global, Local, and Contagious Investor Sentiment.” *Journal of Financial Economics* 104 (2): 272–287.
- Barberis, Nicholas, Andrei Shleifer, and Robert Vishny (1998). “A Model of Investor Sentiment.” *Journal of Financial Economics* 49 (3): 307–343.
- Barberis, Nicholas, and Richard Thaler (2003). “A Survey of Behavioral Finance.” In *Handbook of the Economics of Finance*, Vol. 1B, edited by George Constantinides, Milton Harris, and René Stulz, 1053–1128. Amsterdam: Elsevier.
- Brown, Gregory W., and Michael T. Cliff (2004). “Investor Sentiment and the Near-Term Stock Market.” *Journal of Empirical Finance* 11 (1): 1–27.
- Brown, Gregory W., and Michael T. Cliff (2005). “Investor Sentiment and Asset Valuation.” *Journal of Business* 78 (2): 405–440.
- Center for Research in Security Prices (CRSP) (2024). “CRSP US Stock Database.” *Wharton Research Data Services (WRDS)*. <https://wrds-www.wharton.upenn.edu/>
- Compustat (2024). “Compustat North America Daily.” *Wharton Research Data Services (WRDS)*. <https://wrds-www.wharton.upenn.edu/>
- Daniel, Kent, and Sheridan Titman (1997). “Evidence on the Characteristics of Cross-Sectional Variation in Stock Returns.” *Journal of Finance* 52 (1): 1–33.

- French, Kenneth R. (2024). "Fama–French Factors and Momentum Data." *Kenneth R. French Data Library*. https://mba.tuck.dartmouth.edu/pages/faculty/ken.french/data_library.html
- Floros, Christos (2008). "Investors' Sentiment in the UK Stock Market." *Studies in Economics and Finance* 25 (1): 3–18.
- Fama, Eugene F., and Kenneth R. French (1993). "Common Risk Factors in the Returns on Stocks and Bonds." *Journal of Financial Economics* 33 (1): 3–56.
- Fama, Eugene F., and Kenneth R. French (1996). "Multifactor Explanations of Asset Pricing Anomalies." *Journal of Finance* 51 (1): 55–84.
- Garcia, Diego (2013). "Sentiment during Recessions." *Journal of Finance* 68 (3): 1267–1300.
- González-Izquierdo, Antonio, Marta Gómez-Puig, and Santiago Sosvilla-Rivero (2023). "Investor Sentiment and Stock Returns in the Eurozone: A Sectoral Approach." *International Review of Financial Analysis* 86: 102664.
- Ibbotson, Roger G., Jody L. Sindelar, and Jay R. Ritter (1994). "The Market's Problems with the Pricing of Initial Public Offerings." *Journal of Applied Corporate Finance* 7 (1): 66–74.
- Jegadeesh, Narasimhan, and Sheridan Titman (1993). "Returns to Buying Winners and Selling Losers: Implications for Stock Market Efficiency." *Journal of Finance* 48 (1): 65–91.
- Lakonishok, Josef, Andrei Shleifer, and Robert Vishny (1994). "Contrarian Investment, Extrapolation, and Risk." *Journal of Finance* 49 (5): 1541–1578.
- MSCI (2024). "MSCI Momentum Indexes Methodology." *MSCI Index Methodology*. <https://www.msci.com/index/methodology/latest/Momentum>
- Ritter, Jay R. (2024). "Initial Public Offerings: Updated Statistics." *Jay Ritter's Website*. <https://site.warrington.ufl.edu/ritter/ipo-data/>
- Schmeling, Maik (2009). "Investor Sentiment and Stock Returns: Some International Evidence." *Journal of Empirical Finance* 16 (3): 394–408.
- Stambaugh, Robert F., Jianfeng Yu, and Yu Yuan (2012). "The Short of It: Investor Sentiment and Anomalies." *Journal of Financial Economics* 104 (2): 288–302.
- Tetlock, Paul C. (2007). "Giving Content to Investor Sentiment: The Role of Media in the Stock Market." *Journal of Finance* 62 (3): 1139–1168.

Thaler, Richard H. (1985). “Mental Accounting and Consumer Choice.” *Marketing Science* 4 (3): 199–214.

Thaler, Richard H. (1993). *Advances in Behavioral Finance*. New York: Russell Sage Foundation.

Thaler, Richard H., and H. M. Shefrin (1981). “An Economic Theory of Self-Control.” *Journal of Political Economy* 89 (2): 392–406.

Verma, Rahul, and Priti Verma (2007). “Noise Trading and Stock Market Volatility.” *Journal of Multinational Financial Management* 17 (3): 231–243.

APPENDIX I: DATA CONSTRUCTION

In the following the Data collecting and preparation process is described.

I. Data Collection and Preparation

I.1 Sentiment Data– monthly

Retrieved from: NYU Stern - Baker and Wurgler's Sentiment Index

Variables	3
	Index Standardized Index Sentiment classification

Data Preparation

Excel:

- Raw monthly sentiment index standardized using a z-score transformation based on the full sample mean and standard deviation.
- Months classified as high sentiment when the standardized index is above zero and low sentiment when below zero.
- Standardized index and sentiment classification stored in the Excel file prior to import.

Stata:

- Finalized sentiment data imported from Excel.
- Sentiment variables renamed for consistency.
- Binary high-sentiment indicator created from the imported classification.
- Standardized sentiment index used directly in all analyses without re-standardization.

CRSP – monthly Data

Retrieved from: CRSP US Stock Database Wharton Research Data Services (WRDS)

Raw Dataset

Observations:	2,178,761
Variables	PERMNO Names Date Exchange Code Standard Industrial Classification Code Ticker Symbol PERMCO Delisting Return Price or Bid/Ask Average Returns Shares Outstanding Returns without Dividends

Data Preparation

- Monthly CRSP data were loaded and key variables were renamed for consistency across datasets.
- The sample was restricted to firms listed on NYSE, AMEX, or NASDAQ.
- A monthly time variable was constructed and the sample period was limited to 1999–2023.
- Firm size was computed as market equity using price and shares outstanding.
- Monthly returns were cleaned by excluding invalid observations and adjusted for delisting returns.
- Log returns were constructed to compute return volatility.
- Calendar years were extracted from the date variable.
- Monthly firm-level data were aggregated to annual averages, with return volatility measured as the standard deviation of log returns within each year.
- The cleaned annual CRSP dataset was saved for subsequent merging with accounting and sentiment data.

Prepared Dataset

Observations:	171,088
Variables	(mean) me (mean) adj_ret (sd) logret Year permno

Although CRSP returns are observed at the monthly frequency, firm-level characteristics such as market equity and return volatility are constructed as annual measures based on monthly data. These annual characteristics are then matched to monthly returns in subsequent regressions.

I.2 CRSP – Firm Age Construction (Listed Firm Age)

Retrieved from: CRSP US Stock Database Wharton Research Data Services (WRDS)

Raw Dataset

Observations	5,158,538
Variables	PERMNO Names Date Exchange Code Ticker Symbol

Data Preparation

- Loaded CRSP trading-date data and restricted the sample to NYSE, AMEX, and NASDAQ listings.
- Verified that firm identifiers (PERMNO) and dates were non-missing.
- For each firm, identified the first CRSP observation date as the starting point for the listing-age measure.

- Computed firm age in months and converted it to years for each observation.
- Created calendar year and month variables and flagged observations in December 2023.
- Defined each firm's final age as the December 2023 age if available; otherwise, used the age at the firm's last available CRSP observation.
- Collapsed to one observation per firm containing its final listing age and saved the resulting file for merging with the main dataset.

Prepared Dataset

Observations	34,752
Variables	PERMNO Final Age (Listed Age)

I.3 Compustat Accounting Data – quarterly

Retrieved from: Compustat North America Wharton Research Data Services (WRDS)

Raw Dataset

Observations	985,800
	(costat) Active/Inactive Status Marker (curcdq) ISO Currency Code (datafmt) Data Format (indfmt) Industry Format (consol) Level of Consolidation - Company Interim Descriptor (gvkey) Global Company Key (datadate) Data Date (conm) Company Name (tic) Ticker Symbol (exchg) Stock Exchange Code (fyr) Fiscal Year-end Month (fic) Current ISO Country Code - Incorporation (fyrc) Current Fiscal Year End Month (sic) Standard Industry Classification Code (datacqtr) Calendar Data Year and Quarter (datafqtr) Fiscal Data Year and Quarter (fqtr) Fiscal Quarter (rdq) Report Date of Quarterly Earnings (atq) Assets - Total (ceqq) Common/Ordinary Equity - Total (cshoq) Common Shares Outstanding (cstkq) Common/Ordinary Stock (Capital) (dltdq) Long-Term Debt - Total (dvpq) Dividends - Preferred/Preference (ibq) Income Before Extraordinary Items (ltq) Liabilities - Total (niq) Net Income (Loss) (oibdpq) Operating Income Before Depreciation - Quarterly

	(pstkrq) Preferred/Preference Stock - Redeemable (saleq) Sales/Turnover (Net) (teqq) Stockholders Equity - Total (txdbq) Deferred Taxes - Balance Sheet (txditcq) Deferred Taxes and Investment Tax Credit (uceqq) Common Equity - Total - Utility (xrdq) Research and Development Expense (capxy) Capital Expenditures (oibdpq) Operating Income Before Depreciation (prccq) Price Close - Quarter
--	---

Data Preparation

- Loaded quarterly Compustat fundamentals and restricted the sample to U.S.-incorporated firms.
- Converted accounting dates into a quarterly time format.
- Constructed book equity following the standard Compustat definition.
- Computed firm-level accounting variables, including return on assets, leverage, capital expenditures, sales growth, R&D intensity, and dividend and profitability indicators.
- Removed observations with missing key accounting information.
- Eliminated duplicate firm–quarter observations to retain one record per firm and quarter.
- Standardized variable names to ensure consistency with CRSP and other datasets.
- Saved the cleaned quarterly Compustat dataset for subsequent merging with return and sentiment data.

Prepared Data

Observations	538,826
	(costat) Active/Inactive Status Marker (curedq) ISO Currency Code (datafmt) Data Format (indfmt) Industry Format (consol) Level of Consolidation - Company Interim Descriptor (gvkey) Global Company Key (datadate) Data Date (conm) Company Name (tic) Ticker Symbol (exchg) Stock Exchange Code (fy) Fiscal Year-end Month (fic) Current ISO Country Code - Incorporation (fyrc) Current Fiscal Year End Month (sic) Standard Industry Classification Code (datacqtr) Calendar Data Year and Quarter (datafqtr) Fiscal Data Year and Quarter (fqtr) Fiscal Quarter (rdq) Report Date of Quarterly Earnings (atq) Assets - Total

	(ceqq) Common/Ordinary Equity - Total (cshoq) Common Shares Outstanding (cstkq) Common/Ordinary Stock (Capital) (dlittq) Long-Term Debt - Total (dvpq) Dividends - Preferred/Preference (ibq) Income Before Extraordinary Items (ltq) Liabilities - Total (niq) Net Income (Loss) (oibdpq) Operating Income Before Depreciation - Quarterly (pstkrq) Preferred/Preference Stock - Redeemable (saleq) Sales/Turnover (Net) (teqq) Stockholders Equity - Total (txdbq) Deferred Taxes - Balance Sheet (txditcq) Deferred Taxes and Investment Tax Credit (uceqq) Common Equity - Total - Utility (xrdq) Research and Development Expense (capxy) Capital Expenditures (oibdpy) Operating Income Before Depreciation (prccq) Price Close - Quarter qdate Book Equity (CEQ + [TXDB] + TXDITC - PSTK) Return on Assets (NI / AT) CapEx to Assets Leverage (LT / AT) l_saleq Sales Growth R&D / Assets Dividend Dummy (1=Dividend, 0=No Dividend) Profitability Dummy (1=Profit, 0=Loss)
--	--

I.4 Linktable – CCMX

Retrieved from: Wharton Research Data Services (WRDS)

Raw Data

Observations	32,378
Variables	Global Company Key Ticker Symbol CUSIP Primary Link Marker Security-level Identifier Link Type Codd Historical CRSP PERMNO Link to COMPUSTAT Record Historical CRSP PERMCO Link to COMPUSTAT Record First Effective Date of Link Last Effective Date of Link

- The CRSP–Compustat link table was used to merge firm-level return data from CRSP with accounting information from Compustat.
- Only valid and active link types were retained to ensure accurate firm identification across databases.
- The link table ensured consistent matching of CRSP PERMNO identifiers with Compustat GVKEY identifiers throughout the sample period.

I.5 Risk-Factors

Retrieved from: Kenneth R. French Data Library

- Fama–French risk factors were downloaded to control for common sources of systematic risk in asset returns.
- The momentum factor was included to account for return continuation effects not captured by the standard Fama–French factors.
- These factors were used in multivariate regressions to isolate sentiment-related effects from established risk premia.

II. Data Merging

- CRSP–Compustat link table was used, using primary and active link types to match CRSP PERMNO identifiers with Compustat GVKEYs.
- Monthly CRSP return data were merged with the link table to attach Compustat firm identifiers.
- Monthly CRSP observations were aligned with quarterly Compustat accounting data by mapping calendar months to corresponding fiscal quarters.
- Quarterly accounting variables were carried forward to the monthly frequency within each firm.
- Monthly investor sentiment data were merged by month to the firm-level dataset.
- Firm age information derived from CRSP trading history was merged at the firm level.
- The book-to-market ratio was constructed where necessary, and observations with missing returns were excluded.
- The final merged monthly panel dataset was saved for portfolio construction and regression analysis.

Merged Dataset

Observations:	1,847,597
Variables	PERMNO Names Date Exchange Code Standard Industrial Classification Code Ticker Symbol PERMCO Delisting Return Price or Bid/Ask Average Returns

	Shares Outstanding Returns without Dividends mdate Market Equity in \$1000 logret 12-month rolling volatility of log returns Adjusted return incl. delistingGlobal Company Key fyear year qdate (costat) Active/Inactive Status Marker (curecdq) ISO Currency Code (datafmt) Data Format (indfmt) Industry Format (consol) Level of Consolidation - Company Interim Descriptor (datadate) Data Date (conm) Company Name (tic) Ticker Symbol (exchg) Stock Exchange Code (fyr) Fiscal Year-end Month (fic) Current ISO Country Code - Incorporation (fyrc) Current Fiscal Year End Month (sic) Standard Industry Classification Code (datacqtr) Calendar Data Year and Quarter (datafqtr) Fiscal Data Year and Quarter (fqtr) Fiscal Quarter (rdq) Report Date of Quarterly Earnings (atq) Assets - Total (ceqq) Common/Ordinary Equity - Total (cshoq) Common Shares Outstanding (cstkq) Common/Ordinary Stock (Capital) (dlttq) Long-Term Debt - Total (dvpq) Dividends - Preferred/Preference (ibq) Income Before Extraordinary Items (ltq) Liabilities - Total (niq) Net Income (Loss) (oibdpq) Operating Income Before Depreciation - Quarterly (pstkrq) Preferred/Preference Stock - Redeemable (saleq) Sales/Turnover (Net) (teqq) Stockholders Equity - Total (txdbq) Deferred Taxes - Balance Sheet (txditcq) Deferred Taxes and Investment Tax Credit (uceqq) Common Equity - Total - Utility (xrdq) Research and Development Expense (capxy) Capital Expenditures (oibdpy) Operating Income Before Depreciation (prccq) Price Close - Quarter Book Equity (CEQ + [TXDB] + TXDITC - PSTK) Return on Assets (NI / AT) CapEx to Assets
--	--

	Leverage (LT / AT) l_saleq Sales Growth R&D / Assets Dividend Dummy (1=Dividend, 0=No Dividend) Profitability Dummy (1=Profit, 0=Loss) sentiment_index z_index st_index sentiment_classification high_sentiment float Firm Age in years (at Dec 2023 or last obs) Book-to-Market
--	---

Table A1

Summary Statistics of the Merged Dataset (Before Cleaning)

(707,866 missing observations deleted)

These statistics report firm-month observations prior to the application of sample restrictions and winsorization.

	count	mean	sd	min	p25	p50	p75	max
Market Equity	1123928	5078.004	32406.43	.08355	91.9074	444.5847	2026.829	3071345
Volatility	1123928	.1482554	.1040792	0	.0772861	.1190983	.1874543	3.623404
Return on Assets	1123928	-.027027	.9791211	-341.4445	-.018567	.004885	.0171735	147.035
Adjusted Return	1123928	.0090831	.2032448	-.9958333	-.073494	.0016129	.0739166	17.60465
CapEx / Assets	1123928	.027052	.046758	-.3125	.0033933	.0124722	.0318822	4.045526
Listed Firm Age (years)	1123928	25.35711	18.65182	.1666667	11.08333	21.5	33.08333	98.91666

Source: Author's calculation based on data from CRSP, Compustat, and the Baker and Wurgler Sentiment Index.

Table A2**Summary Statistics of the Merged Dataset by Sentiment (Before Cleaning)**

The table reports descriptive statistics separately for low- and high-sentiment periods as classified by the sentiment indicator.

Low Sentiment	count	mean	sd	min	max
Market Equity	731296	5107.001	29006.76	.088335	3071345
Volatility	731296	.1382843	.0934908	1.63e-06	2.342982
Return on Assets	731296	-.0209176	1.060819	-341.4445	147.035
Adjusted Return	731296	.0122529	.1899709	-.9936	15.98446
CapEx/Assets	731296	.026482	.0460938	-.1857681	4.045526
Listed Firm Age (years)	731296	26.25113	18.6431	.1666667	98.91666
High Sentiment					
Market Equity	392632	5023.997	37934.78	.08355	2902368
Volatility	392632	.1668271	.1191586	0	3.623404
Return on Assets	392632	-.038406	.8050278	-235.5948	21.66499
Adjusted Return	392632	.0031793	.2257799	-.9958333	17.60465
CapEx/Assets	392632	.0281137	.0479527	-.3125	1.409091
Listed Firm Age (years)	392632	23.69194	18.55359	.25	98.91666

Source: Author's calculation based on data from CRSP, Compustat, and the Baker and Wurgler Sentiment Index.

Cleaning of the merged Dataset

- The merged monthly firm-level panel was loaded and declared as a monthly panel structure (PERMNO $\times$ month).
- Firms were required to have a minimum return history of at least 9 out of the previous 12 months to ensure reliable momentum and volatility measures.
- Momentum measures (12–2 and 6–2) and rolling return volatility were constructed using rolling windows and combined into an iShares-style momentum indicator.
- Observations with missing or non-positive values for key inputs (market equity, book equity, firm age, capital expenditures, volatility) were excluded.
- Book-to-market ratios were constructed using valid positive book equity and market equity values.
- Scale variables were log-transformed where economically meaningful and positive.
- Variable-specific winsorization was applied on a year-by-year basis to mitigate the influence of outliers, with cutoffs tailored to the distribution of each variable.
- The resulting dataset was saved as a cleaned, analysis-ready panel for portfolio construction and regression analysis

APPENDIX II: PORTFOLIO CONSTRUCTION DETAILS

Table A3

Portfolio Decile Returns – Excerpt from Variable Size

This table provides an illustrative excerpt of monthly equal-weighted decile returns for the size sort in post-earnings months. The main analysis uses these decile returns to construct Top3–Bottom3 spreads, which are then analysed in the spread regressions reported in Section 4.

mdate	ret1	ret2	ret3	ret4	ret5	ret6	ret7	ret8	ret9	ret10
01.08.2000	-0.13									
01.09.2000	-0.07	-0.10	-0.08	0.00	-0.01	-0.02	0.02	0.02	-0.05	-0.03
01.12.2000	-0.19	-0.04	-0.12	0.01	0.02	0.06	0.07	0.09	0.02	0.07
01.03.2001	-0.12	-0.09	-0.07	-0.08	0.04	-0.07	0.04	-0.06	-0.09	-0.09
01.06.2001	0.00	0.02	0.01	0.00	0.09	0.03	0.01	0.09	0.03	-0.01
01.09.2001	-0.19	-0.17	-0.13	-0.16	-0.18	-0.20	-0.19	-0.24	-0.21	-0.21
01.12.2001	-0.02	0.11	0.08	0.16	0.09	0.10	0.08	0.06	0.03	0.06
01.03.2002	0.07	0.08	0.09	0.08	0.13	0.08	0.05	0.12	0.09	0.06
01.06.2002	-0.13	-0.06	-0.06	-0.06	0.01	-0.04	0.02	-0.06	-0.13	-0.04
01.09.2002	-0.12	-0.15	-0.07	-0.12	-0.11	-0.05	-0.08	-0.10	-0.11	-0.10
01.12.2002	-0.04	-0.15	-0.06	-0.14	-0.06	-0.12	-0.09	-0.12	-0.11	-0.10
01.03.2003	-0.04	0.02	0.07	0.07	0.01	0.02	0.02	0.03	0.02	0.01
01.06.2003	0.03	0.04	0.05	0.02	0.05	0.05	0.02	0.03	0.01	0.00
01.09.2003	0.05	-0.01	0.04	0.02	-0.01	-0.06	-0.03	-0.05	-0.06	-0.06
01.12.2003	-0.16	-0.01	0.00	-0.03	0.02	0.02	-0.02	-0.06	-0.01	0.00
01.03.2004	-0.04	-0.01	-0.04	-0.03	0.07	0.01	0.01	0.01	0.03	0.01
01.06.2004	-0.02	-0.02	0.12	0.07	0.03	0.05	0.04	0.05	0.04	0.01
01.09.2004	0.02	0.08	0.01	0.05	0.07	0.09	0.05	0.05	0.04	0.05
01.12.2004	0.09	0.05	0.09	0.06	0.04	0.05	0.05	0.10	0.04	0.04
01.03.2005	-0.05	-0.10	-0.07	0.03	-0.05	-0.02	0.00	0.02	0.04	-0.01
01.06.2005	0.04	0.01	0.06	0.11	0.06	0.01	0.04	0.01	0.02	0.01
01.09.2005	0.03	0.01	-0.03	-0.02	0.01	-0.01	-0.04	0.02	0.00	-0.03
01.12.2005	-0.01	0.02	-0.01	-0.02	0.00	0.00	0.00	0.00	0.00	0.00
01.03.2006	0.06	0.09	0.02	0.03	0.03	0.05	0.03	0.05	0.03	0.00
01.06.2006	-0.03	0.00	-0.02	-0.03	-0.01	-0.04	0.01	-0.03	-0.01	0.01
01.09.2006	0.00	0.05	0.07	0.07	0.05	0.08	0.09	0.06	0.06	0.07
01.12.2006	0.00	0.02	0.04	-0.01	-0.01	-0.02	-0.02	-0.02	0.01	0.00
01.03.2007	-0.04	0.03	0.00	-0.02	0.00	0.04	0.06	-0.03	-0.02	-0.02
01.06.2007	0.04	-0.01	0.01	-0.07	0.00	-0.04	-0.03	-0.03	0.00	-0.01
01.09.2007	-0.04	0.02	-0.06	-0.04	-0.02	0.01	0.00	0.00	0.04	-0.03
01.12.2007	-0.03	0.02	-0.01	-0.06	-0.07	-0.03	-0.02	0.00	0.00	-0.02
01.03.2008	-0.01	0.04	0.00	0.02	-0.01	-0.01	0.01	-0.03	0.01	0.00
01.06.2008	-0.18	-0.12	-0.12	-0.13	-0.08	-0.10	-0.09	-0.09	-0.12	-0.07
01.09.2008	-0.16	-0.16	-0.06	-0.04	-0.11	-0.10	-0.08	-0.07	-0.14	-0.08
01.12.2008	-0.18	-0.09	0.30	0.25	0.03	0.14	0.05	0.08	0.01	0.03
01.03.2009	-0.01	0.29	0.04	0.05	0.16	0.19	0.19	0.14	0.12	0.15

Source: Author's calculation based on data from CRSP, Compustat, and the Baker and Wurgler Sentiment Index.

APPENDIX III: ROBUSTNESS CHECKS

III. 1 Spread Regressions with High-Sentiment Dummy

Those tables report spread regressions using a high-sentiment dummy, defined as an indicator equal to one in months classified as high sentiment and zero otherwise, and simple bootstrap standard errors (200 replications) to assess the robustness of statistical inference. The results are fully consistent with the main findings: size-related spreads remain positive and statistically significant, while book-to-market and volatility effects weaken once risk factors are controlled for. Overall, the bootstrap inference confirms that the core conclusions do not depend on asymptotic standard error assumptions.

Table A4
Univariate Regression of characteristic spreads (sentiment dummy)

	Size	BM	ROA	Listed Age	CapEx/Assets	Volatility	Mom	Mom _{msci}
Sentiment	0.0233**	-0.0169**	0.0122	0.0154*	-0.0026	-0.0266*	0.0148	0.0205*
(SE)	(0.0113)	(0.0085)	(0.0099)	(0.0086)	(0.0087)	(0.0156)	(0.0115)	(0.0110)
_cons	-0.0048	-0.0151***	0.0027	-0.0076	-0.0003	0.0019	-0.0067	-0.0055
(SE)	(0.0063)	(0.0058)	(0.0061)	(0.0052)	(0.0047)	(0.0107)	(0.0051)	(0.0050)
Obs.	94	94	94	94	94	94	93	93
R-squared	0.047	0.033	0.017	0.034	0.001	0.028	0.022	0.043

Source: Author's calculation based on data from CRSP, Compustat, and the Baker and Wurgler Sentiment Index.

Table A5
Multivariate Regression of characteristic spreads (sentiment dummy)

	Size	BM	ROA	Listed Age	CapEx/Assets	Volatility	Mom	Mom _{msci}
Sentiment	0.0193*	-0.0113	0.0045	0.0068	0.0007	-0.0070	0.0101	0.0134
	(0.0102)	(0.0082)	(0.0084)	(0.0085)	(0.0083)	(0.0104)	(0.0120)	(0.0103)
rmrf	-0.0033**	0.0022*	-0.0027**	-0.0019*	-0.0025*	0.0102***	-0.0033**	-0.0050***
	(0.0015)	(0.0013)	(0.0012)	(0.0011)	(0.0013)	(0.0015)	(0.0016)	(0.0013)
hml	-0.0052**		0.0009	0.0035**	0.0013	0.0023	-0.0020	-0.0031
	(0.0025)		(0.0019)	(0.0017)	(0.0019)	(0.0029)	(0.0023)	(0.0021)
rmw	-0.0011	0.0035	0.0026	0.0025	0.0041*	-0.0017	0.0023	0.0026
	(0.0028)	(0.0028)	(0.0027)	(0.0021)	(0.0021)	(0.0031)	(0.0032)	(0.0029)
cma	0.0062	0.0016	-0.0043	0.0007	-0.0049*	-0.0040	0.0001	0.0009
	(0.0040)	(0.0025)	(0.0026)	(0.0023)	(0.0030)	(0.0044)	(0.0042)	(0.0040)
umd	-0.0011	-0.0029*	0.0027	0.0012	-0.0033**	-0.0012		
	(0.0020)	(0.0016)	(0.0018)	(0.0012)	(0.0017)	(0.0020)		
smb		0.0033*	-0.0017	-0.0024	0.0013	0.0059***	0.0003	0.0009
		(0.0019)	(0.0017)	(0.0016)	(0.0022)	(0.0021)	(0.0022)	(0.0022)
_cons	-0.0030	-0.0186***	0.0034	-0.0068	-0.0001	-0.0049	-0.0063	-0.0051
	(0.0057)	(0.0057)	(0.0053)	(0.0050)	(0.0050)	(0.0062)	(0.0048)	(0.0039)
Obs.	94	94	94	94	94	94	93	93
R-squared	0.213	0.261	0.310	0.261	0.154	0.648	0.182	0.365

Standard errors in parentheses Simple Bootstrap (200 Reps), inner vce(ols); Eventmonths

* $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

Source: Author's calculation based on data from CRSP, Compustat, and the Baker and Wurgler Sentiment Index.

III.2 Firm-Level Panel Regressions: Fixed Effects Comparison

Table A6

Firm-Level Panel Regressions with Bootstrap Standard Errors

This table reports firm-level panel regressions with bootstrap standard errors clustered at the firm level (200 replications). The results closely mirror the main fixed effects estimates: interaction effects between sentiment and firm characteristics, most notably size and book to market, remain statistically significant, while the baseline sentiment coefficient becomes insignificant once fixed effects are included. Overall, the bootstrap inference confirms that the main conclusions are not driven by asymptotic standard error assumptions.

Firm-level panel: 1M-lag sentiment – FE comparison (bootstrap SEs)

	(1) adj ret	(2) adj ret	(3) adj ret
Sentiment _{t-1}	-0.0751*** (0.0223)	-0.0275 (0.0254)	-0.0283 (0.0256)
BM _{t-1}	0.0067** (0.0026)	0.0120*** (0.0045)	0.0111** (0.0044)
Size _{t-1}	-0.0010 (0.0010)	-0.0110*** (0.0033)	-0.0112*** (0.0034)
ROA _{t-1}	0.3304*** (0.0879)	0.3235*** (0.0913)	0.3289*** (0.0915)
Volatility _{t-1}	0.0020 (0.0045)	0.0018 (0.0052)	0.0017 (0.0053)
Momentum _{t-1}	-0.0022 (0.0048)	0.0011 (0.0049)	0.0010 (0.0050)
Listed Age _{t-1}	-0.0026 (0.0030)	-4.7958 (3.4850)	-8.4196*** (2.7639)
Sentiment x BM	-0.0120*** (0.0046)	-0.0134*** (0.0043)	-0.0130*** (0.0043)
Sentiment x Size	-0.0040** (0.0020)	-0.0039* (0.0022)	-0.0039* (0.0022)
Sentiment x ROA	-0.2753** (0.1230)	-0.3059** (0.1277)	-0.3053** (0.1279)
Sentiment x Volatility	-0.0257*** (0.0085)	-0.0207** (0.0083)	-0.0205** (0.0084)
Sentiment x Momentum	-0.0023 (0.0082)	-0.0057 (0.0090)	-0.0055 (0.0090)
Sentiment x Listed Age	0.0091* (0.0050)	-0.0004 (0.0071)	0.0001 (0.0071)
N	9110	9056	9055

Standard errors in parentheses

Bootstrap SEs (200 reps), clustered by firm (permno)

* $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

Source: Author's calculation based on data from CRSP, Compustat, and the Baker and Wurgler Sentiment Index.