Band 120

Lukas Lautenschläger

Behavioral Finance in Real Estate: Textual Sentiment and REIT Market Dynamics

IRE|BS
International Real Estate Business School
Universität Regensburg

Schriften zu Immobilienökonomie und Immobilienrecht

Herausgeber:
IREIBS International Real Estate Business School

Prof. Dr. Sven Bienert
Prof. Dr. Stephan Bone-Winkel
Prof. Dr. Kristof Dascher
Prof. Dr. Dr. Herbert Grziwotz
Prof. Dr. Tobias Just
Prof. Gabriel Lee, Ph. D.
Prof. Dr. Kurt Klein
Prof. Dr. Jürgen Kühling, LL.M.
Prof. Dr. Gerrit Manssen
Prof. Dr. Dr. h.c. Joachim Möller
Prof. Dr. Karl-Werner Schulte HonRICS
Prof. Dr. Wolfgang Schäfers
Prof. Dr. Steffen Sebastian
Prof. Dr. Wolfgang Servatius
Prof. Dr. Frank Stellmann
Prof. Dr. Martin Wentz

Lukas Lautenschläger

Behavioral Finance in Real Estate: Textual Sentiment and REIT Market Dynamics

Die Deutsche Bibliothek – CIP Einheitsaufnahme
Lukas Lautenschläger
Behavioral Finance in Real Estate: Textual Sentiment and REIT Market Dynamics
Regensburg: Universitätsbibliothek Regensburg 2026
(Schriften zu Immobilienökonomie und Immobilienrecht; Bd. 120)
Zugl.: Regensburg, Univ. Regensburg, Diss., 2026
ISBN 978-3-88246-566-2

ISBN 978-3-88246-566-2
© IRE|BS International Real Estate Business School, Universität Regensburg
Verlag: Universitätsbibliothek Regensburg, Regensburg 2026
Zugleich: Dissertation zur Erlangung des Grades eines Doktors der Wirtschaftswissenschaften, eingereicht an der Fakultät für Wirtschaftswissenschaften der Universität Regensburg
Tag der mündlichen Prüfung: 29. Juli 2026
Berichterstatter: Prof. Dr. Wolfgang Schäfers
Prof. Dr. Tobias Just

Declaration on the Use of Artificial Intelligence (AI)

In the preparation of this dissertation, AI-assisted tools and large language models were used to improve spelling and grammar, and to support the linguistic optimization of my own wording. After using these tools, the content was independently reviewed and, where necessary, revised. I take full responsibility for the content of the submitted academic work. As of the time of writing, no official or specific guidelines of the Faculty of Business, Economics, and Real Estate of the University of Regensburg regarding the use of AI in dissertations have been adopted. Therefore, reference is made to the guidelines of the Faculty of Law of the University of Regensburg, dated May 7, 2025.

Table of Contents

List of Tables ..... **VII**

List of Figures ..... **VIII**

1 Introduction **1**

1.1 Motivation and Background 1

1.2 Research Questions 6

1.3 Co-Authors, Submissions and Conference Presentations 10

1.4 References 14

2 Social Media and Real Estate: Do Twitter Users Predict REIT Performance? **17**

2.1 Abstract 17

2.2 Introduction 18

2.3 Literature Review and Hypothesis Development 19

2.3.1 Sentiment in Real Estate 19

2.3.2 Social Media Sentiment in Finance 22

2.3.3 Complexity of Social Media Data 23

2.3.4 Hypotheses 24

2.4 Data 25

2.4.1 Social Media Data 25

2.4.2 Dependent Variable 26

2.4.3 Economic Control Variables 27

2.5 Methodology 27

2.5.1 Tweet Selection and Data Cleaning 28

2.5.2 Sentiment Analysis 30

2.5.3 Econometric Approaches 35

2.6 Results 36

2.6.1 Tweet Selection 36

2.6.2 Model Accuracy 37

2.6.3 Sentiment Indices 37

2.6.4 Vector Autoregression 39

2.6.5 Granger Causality and Impulse Response Functions 42

2.7 Conclusion 44

2.8 Appendix 46

2.8.1 Appendix I – Support Vector Machine 46

2.8.2 Appendix II – Long Short-Term Memory 47

2.8.3 Appendix III – Signal Words ... 48
2.8.4 Appendix IV – Vector Autoregression: Quarterly Lags ... 49
2.9 References ... 51
3 Fundamentals or Noise? The Informative Value of Twitter Sentiment for REIT Returns ... 59
3.1 Abstract ... 59
3.2 Introduction ... 60
3.3 Literature Review ... 63
3.3.1 Individual Investor Behavior ... 63
3.3.2 Individual Investors and Social Media ... 64
3.3.3 Twitter Sentiment and Stock Market Fundamentals ... 65
3.4 Data and Methodology ... 66
3.4.1 Derivation of Twitter Sentiment Measures ... 66
3.4.2 Macroeconomic, Capital and Real Estate Market Variables ... 69
3.4.3 Estimation ... 70
3.5 Results ... 72
3.5.1 Twitter Sentiment and Fundamental Variables ... 72
3.5.2 The Relationships of Rational and Irrational Twitter Sentiment and REIT Returns ... 76
3.6 Conclusion ... 78
3.7 References ... 81
4 Signals or Stories? ESG Sentiment and Investor Reactions in REIT Corporate Disclosure ... 87
4.1 Abstract ... 87
4.2 Introduction ... 88
4.3 Literature Review ... 90
4.3.1 Overall Effect of ESG in Corporate Disclosure ... 90
4.3.2 Sentiment Analysis as a Tool to Assess ESG Performance ... 91
4.3.3 A Disaggregated View of the ESG Impact ... 92
4.3.4 Heterogeneity of ESG Sentiment in REIT Subsectors ... 93
4.4 Data and Methodology ... 94
4.4.1 Filtering for ESG Relevance ... 94
4.4.2 Derivation of Tone Measures ... 95
4.4.3 Event Study and Regression Analysis ... 95
4.5 Results ... 97

4.6 Conclusion 106
4.7 References 107
5 News Sentiment and Stock Return Predictability: The Case of U.S. REITs 113
5.1 Abstract 113
5.2 Introduction 114
5.3 Related Literature and Hypotheses Development 117
5.3.1 Return Predictability and Relative Sector Performance 117
5.3.2 Large Language Models and Textual Analysis 118
5.4 Data 120
5.4.1 Capital Market and Macroeconomic Data 120
5.4.2 News and Sentiment Data 121
5.5 Methodology 122
5.5.1 Sentiment Classification with LLMs 122
5.5.2 Construction of Sentiment Indices 123
5.5.3 Forecasting Framework 124
5.5.4 Model Description and Evaluation Metrics 125
5.6 Results 128
5.6.1 Regression Analysis 128
5.6.2 Directional Movement Prediction 129
5.6.3 Directional Excess Return Prediction 131
5.6.4 Long Signal Trading Strategy 135
5.7 Conclusion 138
5.8 Appendix 140
5.8.1 Appendix I – SentFiN 1.0 Prompt 140
5.8.2 Appendix II – Sentiment Classification Prompt 141
5.8.3 Appendix III – Mitigation of GPT's Look-Ahead Bias 142
5.8.4 Appendix IV – Recall Accuracy Prompts 145
5.9 References 146
6 Conclusion 155
6.1 Executive Summary 155
6.2 Final Remarks 165
6.3 References 168

List of Tables

Table 2.1 – Descriptive Statistics 26
Table 2.2 – VAR Results: Monthly Lags Optimism Indicator 39
Table 2.3 – VAR Results: Monthly Lags Pessimism Indicator 40
Table 2.4 – VAR Results: Monthly Lags Sentiment Quotient 41
Table 2.5 – Signal Words 48
Table 2.6 – VAR Results: Quarterly Lags Optimism Indicator 49
Table 2.7 – VAR Results: Quarterly Lags Pessimism Indicator 49
Table 2.8 – VAR Results: Quarterly Lags Sentiment Quotient 50
Table 3.1 – Descriptive Statistics 70
Table 3.2 – ARDL Results: Sentiment, Inflation, and Unemployment Rate 73
Table 3.3 – ARDL Results: Sentiment, S&P 500 Returns and Credit Risk Spread 74
Table 3.4 – ARDL Results: Sentiment, Stock Market Sentiment, and Stock Market Volatility 75
Table 3.5 – ARDL Results: Sentiment and Commercial Real Estate Prices 76
Table 3.6 – VAR Results: Rational and Irrational Sentiment and REIT Returns (Pre-COVID) 77
Table 3.7 – VAR Results: Rational and Irrational Sentiment and REIT Returns (COVID) 77
Table 4.1 – Descriptive Statistics 97
Table 4.2 – Regression Results: Overall and Overall ESG Sentiment 98
Table 4.3 – Regression Results: Pillar-Level ESG Sentiment 99
Table 4.4 – Regression Results: Environmental and Sector-Level Sentiment 102
Table 4.5 – Regression Results: Social and Sector-Level Sentiment 103
Table 4.6 – Regression Results: Governance and Sector-Level Sentiment 104
Table 4.7 – Regression Results: Overall ESG and Sector-Level Sentiment 105
Table 5.1 – Descriptive Statistics 120
Table 5.2 – Descriptive Statistics of Weekly Articles Counts 122
Table 5.3 – Performance Metrics for Sentiment Classification on the SEntFiN 1.0 Dataset 123
Table 5.4 – Results: Return Prediction 129
Table 5.5 – Results: Directional Return Prediction 130
Table 5.6 – Results: Directional Excess Return Prediction 132
Table 5.7 – Results: Top-1 Long Strategy 135
Table 5.8 – Results: Top-2 Long Strategy 138
Table 5.9 – Evaluation of Date Recall Accuracy for Historical News Articles 143
Table 5.10 – Evaluation of Recall Accuracy for FTSE EPRA Nareit REIT Indices 144

List of Figures

Figure 2.1 – Unfiltered Tweet Distribution over Time 26
Figure 2.2 – NLP Process 27
Figure 2.3 – LSTM Model 33
Figure 2.4 – Filtered Tweet Distribution over Time 37
Figure 2.5 – Optimism Indicators and Monthly Total Return over Time 38
Figure 2.6 – Pessimism Indicators and Monthly Total Return over Time 38
Figure 2.7 – Sentiment Quotients and Monthly Total Return over Time 38
Figure 2.8 – Impulse Response Functions 43
Figure 3.1 – REIT-Specific Twitter Message Volume per Month 68
Figure 4.1 – Share of ESG Paragraphs in the ESG Reports and Form 10-Ks/10-Qs 95
Figure 5.1 – Out-of-Sample SHAP Absolute Feature Importance 134
Figure 5.2 – Top-1 Long Trading Strategy 137

1 Introduction

1.1 Motivation and Background

In traditional financial theory, capital markets are assumed to be informationally efficient. This assumption is formalized by Fama (1970, p. 383) with the efficient market hypothesis, according to which prices "fully reflect" all available information and, consequently, systematic deviations from fundamental values should not persist. Grossman and Stiglitz (1980), however, argue that full informational efficiency is itself logically inconsistent when information acquisition is costly, because investors who bear the costs would receive no compensation if prices already contained such information. As a consequence, markets cannot be fully efficient but operate in an equilibrium of partial information revelation. The insufficiency of the efficient market framework lies not only in its logical inconsistency but also in its assumption about investor rationality. Drawing on the insight that individuals exhibit systematic deviations from rational choice (Kahneman & Tversky, 1979), the field of behavioral finance began to challenge the traditional paradigm. The idea of irrationality as a driver of the market is not entirely new but echoes Keynes's (1936, p. 161) early notion of *animal spirits*, described as "a spontaneous urge to action rather than inaction."

Black (1986) translates these considerations into a market-level framework through his concept of *noise.* He argues that noise can take different forms. For instance, it may represent information that has already been priced in or, in line with Keynes (1936), the simple act of engaging in trading because one wants to trade. In this framework, noise is attributed a central role in providing liquidity to the market. Without it, there would be nothing to trade against. To capture this dynamic, he differentiates between *information traders*, who trade on information, and *noise traders*, who act on noise as if it were information. As both trade in the same market, prices reflect a combination of genuine information and noise. Black notes, however, that investors cannot reliably tell whether the signals they are acting on constitute information or merely noise. Building directly on this noise trader premise, Shleifer and Summers (1990) propose an alternative to the efficient market perspective that rests on two major assumptions. First, some investors trade on beliefs not fully justified by fundamentals, and second, arbitrage against such trading is risky and therefore limited.

The arbitrage assumption, in particular, has been developed along two complementary lines. De Long et al. (1990) formalize *noise trader risk* as a distinct source of systematic risk. Even if a mispricing caused by noise traders can be correctly identified, an arbitrageur with a short investment horizon who bets against such mispricing risks that noise trader beliefs intensify and drive prices further from fundamentals before they recover, forcing liquidation at a loss. This

undiversifiable risk deters arbitrage, weakening the market's corrective forces. Shleifer and Vishny (1997) trace the *limits of arbitrage* along a different line, attributing them to the way arbitrage capital is organized. In practice, such trading is concentrated not among a large number of small, diversified traders but among a few specialized investors who combine their expertise with external capital. This separation of knowledge and capital creates an agency problem since capital providers cannot reliably distinguish temporary losses from actual underperformance. When a position moves against the arbitrageur in the short run, capital providers may withdraw funding, tightening capital constraints precisely when the mispricing is largest and the expected return from its correction is highest. In extreme cases, arbitrageurs are forced to liquidate positions rather than expand them, leaving the mispricing uncorrected.

These two mechanisms imply that arbitrage is limited by both the behavior of noise traders themselves (De Long et al., 1990) and the institutional structure through which arbitrage capital is allocated (Shleifer & Vishny, 1997). Prices may therefore persistently diverge from fundamental values, making it economically relevant to understand the non-fundamental beliefs driving such deviations. Accordingly, the central task of asset pricing is no longer to identify the source of risk alone, but to examine how expected returns are shaped jointly by risk and investor misvaluation (Hirshleifer, 2001).

One empirically observable form of these non-fundamental beliefs leading to such misvaluation is *investor sentiment*, which Baker and Wurgler (2007, p. 129) define as "a belief about future cash flows and investment risks that is not justified by the facts at hand." Although their definition centers on beliefs held by investors, the underlying concept extends to evaluative signals in market-relevant communication that either express or shape such beliefs. Importantly, Baker and Wurgler (2006) contend that sentiment should not affect all assets equally and that its impact should be greatest for assets that are hard to value and difficult to arbitrage, as these conditions create uncertainty about fundamental values and limit the correction of mispricing. Building on this argument, Real Estate Investment Trusts (REITs) provide a particularly suitable setting for examining these effects, as their structural characteristics may amplify the role of sentiment.

REITs combine public equity market trading with exposure to direct real estate markets, where high information asymmetry prevails, and the underlying physical assets are idiosyncratic, illiquid, and commonly valued through infrequent, backward-looking appraisals. Gallimore and Gray (2002) argue that these characteristics make direct real estate markets particularly susceptible to sentiment-driven behavior. The hybrid nature of REITs makes them more complex to value than many conventional equities, as investors must combine information from the broad equity market with opaque, asset-specific signals. In addition to the complexity of the underlying

properties, the broader REIT market is characterized by high structural diversity, as its subsectors are tied to different tenant industries and, consequently, to distinct business models. This heterogeneity implies that individual subsectors are exposed to different economic drivers and information flows. Office REITs, for instance, are more directly tied to changes in work environments, while industrial REITs are more exposed to shifts in logistics, supply chains, and global trade. Together, these characteristics make this asset class difficult to value and difficult to arbitrage, thereby meeting the conditions under which Baker and Wurgler (2006) argue that sentiment effects are most pronounced. The U.S. REIT market, having grown into the largest and most mature REIT market globally since regulatory reforms in the early 1990s triggered substantial institutional growth (Bone-Winkel et al., 2008; Oh & Verstein, 2024), provides the long data histories and pronounced variation across specialized subsectors necessary to examine these effects empirically.

Investor sentiment is, however, inherently difficult to quantify (Baker & Wurgler, 2007), leading to the development of various empirical proxies. These differ mainly in what they measure and in how directly they observe investor beliefs. Early research on sentiment in real estate relied heavily on survey-based metrics (Clayton et al., 2009; Ling et al., 2014; Freybote, 2016). These measures capture sentiment by soliciting responses to predefined questions rather than recording what participants would communicate unprompted or reveal through their actions. Therefore, surveys elicit sentiment rather than observing it, which introduces both selection bias (Freybote, 2016) and the risk of strategic misrepresentation (Beracha et al., 2019). Market-based sentiment measures, which infer sentiment indirectly from prices, are also widely used in the literature (e.g., Marcato & Nanda, 2016; Freybote & Seagraves, 2017). However, they also face major limitations as they derive sentiment as a residual of market behavior, capturing not the information itself but its aggregated price effect ex post. Consequently, neither of these measures observes sentiment where it originates and circulates.

A more direct source of sentiment comes from the channels through which beliefs are formed and communicated. Shiller (2017) describes these beliefs as embedded in narratives that spread across market participants and can affect economic decisions. Based on this notion, text provides a natural empirical layer on which such beliefs become observable and is therefore not merely another proxy of sentiment. While measuring sentiment does not capture the full content or structure of a narrative, it reflects its evaluative direction, indicating whether it is seen as favorable or unfavorable. Apart from this conceptual rationale, text also offers practical benefits over traditional sentiment measures. It is domain-specific, published frequently, and not aggregated across investor types or delayed through survey cycles. The growing availability of digital text, combined with rapid advances in computational methods, has significantly expanded the accuracy and scale with which sentiment can be extracted from textual sources.

The idea of textual sentiment analysis was introduced to the broader finance literature by Tetlock (2007), who demonstrated that media pessimism predicts downward pressure on markets using a dictionary-based approach. Later, Loughran and McDonald (2011) showed that general-purpose dictionaries misclassify a substantial share of financial vocabulary, establishing that domain-specific language is essential for accurate textual analysis. An early application to real estate followed with Walker (2014), who applied dictionary-based sentiment extraction to housing market news and demonstrated predictive relationships with price dynamics. Subsequent studies broadened the scope of textual corpora and developed domain-specific dictionaries tailored to real estate language (e.g., Beracha et al., 2019; Carstens & Freybote, 2019; Koelbl, 2020; Ruscheinsky et al., 2018). To overcome the main drawback of dictionary-based approaches, namely their context-blindness, machine learning-based classifiers were introduced to the real estate literature by Hausler et al. (2018) and Braun et al. (2020). More recently, the advent of pre-trained transformer models and large language models has further advanced methodological capabilities by substantially improving contextual understanding and domain-specific language processing in sentiment extraction (Devlin et al., 2019; Huang et al., 2023; Lopez-Lira & Tang, 2023).

Despite this methodological progress, the scope of textual sources examined in the REIT literature has remained narrow. The majority of studies focus on institutional texts, such as financial news or corporate disclosures, while other channels through which beliefs form and circulate, such as social media, remain underexplored. This is particularly relevant given two developments that have created new, structurally distinct channels for sentiment. First, the rise of social media and its integration into everyday information consumption have fundamentally changed how both retail and institutional investors access and process information. Social media platforms provide real-time, mostly unfiltered expressions of beliefs and opinions that may precede and drive price movements in ways that institutional text sources do not capture. Second, the growing regulatory pressure and shifting investor preferences toward sustainability have made language related to environmental, social, and governance (ESG) in corporate communication a source of sentiment in its own right. As institutional investors incorporate ESG criteria into their allocation decisions, sentiment in ESG-related corporate communication may convey value-relevant information that a general analysis of such disclosures would otherwise overlook. The limited coverage of these channels in REIT research leaves their implications largely unaddressed.

Against this background, the overarching aim of this dissertation is to determine whether sentiment derived from distinct textual sources carries value-relevant information that helps explain and predict outcomes in the U.S. REIT market. To this end, the following chapters examine social media, corporate communication, and financial news through four papers that each investigate a different aspect of how sentiment forms, transmits, and affects the REIT

market. The first two focus on social media as an investor-driven channel, examining both its predictive relationship with REIT returns and whether that relationship is driven by fundamental information or noise. The third turns to firm-generated corporate communication and analyzes whether ESG-related sentiment in REIT disclosures is reflected in market reactions across different ESG dimensions and property types. While these three papers establish that sentiment from each source is associated with REIT returns, they leave open the question of whether such signals can be actively exploited. Accordingly, the fourth paper moves from an explanatory to a forecasting perspective, testing whether news-based sentiment can predict REIT market outcomes and improve sector allocation decisions beyond what conventional predictors achieve.

Across four studies, this dissertation draws on behavioral asset pricing theory and recent advances in natural language processing to analyze the value relevance of textual sentiment in light of the specific structure of the REIT market. Its central contribution is to show that sentiment should not be understood as a homogeneous signal but as a channel-specific expression of how beliefs are formed, transmitted, and processed in securitized real estate markets. In doing so, it applies methods ranging from dictionary-based approaches to generative language models and demonstrates that signal quality depends not only on model sophistication but also on domain-specific calibration.

1.2 Research Questions

This section presents the course of analysis and the research questions addressed in the four papers of this dissertation. The papers are linked by the overarching question of whether sentiment extracted from social media, corporate communication, and financial news contains information that helps explain and predict REIT market outcomes. Across three information channels, they progressively deepen the analysis, from establishing the sentiment-return relation and decomposing its drivers to translating sentiment signals into out-of-sample forecasts. The following subsections outline how each paper addresses a specific part of this broader research question.

Paper 1: Social Media and Real Estate: Do Twitter Users Predict REIT Performance?

Paper 1 serves as the empirical starting point of this dissertation and asks whether sentiment on social media carries information relevant to the U.S. REIT market. Social media represents an underexplored source of investor sentiment that differs substantially from corporate communication and professional news, as it captures investor communication in a more immediate and less regulated setting. At the same time, its unstructured nature creates several methodological challenges, as social media messages are often short, informal, and noisy. This paper therefore applies extensive data filtering and text preprocessing before comparing dictionary-based, machine learning, and deep learning sentiment classifiers. It tests whether sentiment expressed on social media is predictive of U.S. REIT market returns and assesses which classification approach best captures sentiment in this setting.

Research Questions:

- Does sentiment expressed in social media affect future REIT market returns?
- Is the sentiment expressed in social media related to short-term and long-term fluctuations in REIT market returns?
- Are more sophisticated sentiment classifiers more accurate in their classification performance?
- Do more sophisticated sentiment classifiers provide better predictions of future REIT market returns?

Paper 2: Fundamentals or Noise? The Informative Value of Twitter Sentiment for REIT Returns

Paper 2 builds upon the predictive relationship between social media sentiment and U.S. REIT market returns examined in Paper 1. It shifts the focus from whether a relationship exists to why such a relationship may arise. To address this question, the paper decomposes social media sentiment into a rational component that reflects market fundamentals, and an irrational component that captures noise. The COVID-19 pandemic provides a structural break in the sample period, as both investor attention to REITs and the underlying real estate sector changed substantially, potentially altering the investor composition. This paper tests whether the predictive power of social media sentiment for U.S. REIT returns is driven by fundamental information or by noise, and whether the informative value changed during the pandemic.

Research Questions:

- Which fundamentals drive REIT-specific social media sentiment, and did their explanatory power differ between the pre-COVID and COVID periods?
- Is the relationship between social media sentiment and REIT returns driven by rational, fundamentals-based sentiment or by irrational sentiment?
- Did the informative value of social media sentiment for REIT returns change between the pre-COVID and COVID periods?

Paper 3: Signals or Stories? ESG Sentiment and Investor Reactions in REIT Corporate Disclosure

Paper 3 shifts the analysis from social media content to firm-generated corporate communication as another channel through which beliefs may be expressed. As institutional investors increasingly incorporate ESG criteria into their allocation decisions, ESG-related language in REIT disclosure may convey value-relevant information that a general analysis would not capture. Drawing on mandatory 10-K and 10-Q filings as well as voluntary ESG reports, this paper applies fine-tuned language models to identify ESG-related text and classify its sentiment. It tests whether ESG-related sentiment in REIT disclosure is reflected in cumulative abnormal returns around disclosure dates across ESG pillars, disclosure channels, and REIT subsectors.

Research Questions:

- Is ESG-related sentiment in REIT corporate disclosure associated with cumulative abnormal returns around disclosure dates?
- Do investor reactions to ESG-related sentiment differ between mandatory filings and voluntary ESG reports?
- Does sentiment regarding individual ESG pillars provide more explanatory power for REIT market reactions than aggregate ESG sentiment?
- Does the relationship between ESG-related sentiment and market reactions vary across REIT subsectors?

Paper 4: News Sentiment and Stock Return Predictability: The Case of U.S. REITs

Paper 4 moves from explaining the sentiment-return relation to an out-of-sample forecasting perspective, where the question becomes whether sentiment can predict returns rather than merely explain them. The paper constructs weekly sentiment measures from REIT-related financial news, macroeconomic and policy news, and news on direct real estate markets, and compares sentiment extracted with a fine-tuned language model to sentiment derived using a generative language model. Using a nonlinear machine learning approach and a linear benchmark model for forecasting, this paper tests whether these sentiment signals improve forecasts of aggregate U.S. REIT returns and relative REIT sector performance, and whether they can be translated into economically exploitable sector allocation decisions.

Research Questions:

- Are weekly REIT index returns predictable at the aggregate index level?
- Does news-based investor sentiment contain economically relevant information for predicting the relative performance of REIT subsectors?
- Do sector-specific sentiment scores derived using a generative language model provide stronger predictive signals for REIT returns than aggregate sentiment derived from a fine-tuned language model?

1.3 Co-Authors, Submissions and Conference Presentations

The following overview provides information about each paper's co-authors, journal submission, publication status, and conference presentations.

Paper 1: Social Media and Real Estate: Do Twitter Users Predict REIT Performance?

Authors:

Dr. Nino Paulus, Lukas Lautenschläger, Prof. Dr. Wolfgang Schäfers

Submission Details:

Journal: Journal of Real Estate Research

Current Status: Accepted (01/20/2024) and published in Volume 47, Issue 3 (09/09/2025)

Conference Presentations:

This paper was presented at:

- the 29th Annual Conference of the European Real Estate Society (ERES) in London, England (2023)
- the 40th Annual Conference of the American Real Estate Society (ARES) in Orlando, USA (2024)

Awards

This paper was awarded the "Best PhD Paper Award 2023" at the 29th Annual Conference of the European Real Estate Society (ERES) in London, England (2023) and the 2023 "Manuscript Prize" in the category "Real Estate Investment Trusts" of the American Real Estate Society (ARES).

Paper 2: Fundamentals or Noise? The Informative Value of Twitter Sentiment for REIT Returns

Authors:

Lukas Lautenschläger, Sophia Bodensteiner, Prof. Julia Freybote (PhD),
Prof. Dr. Wolfgang Schäfers

Submission Details:

Journal: Journal of Property Research

Current Status: Revise and Resubmit (04/30/2026)

Conference Presentations:

This paper was presented at:

- the 30th Annual Conference of the European Real Estate Society (ERES) in Gdańsk, Poland (2024)
- the 41st Annual Conference of the American Real Estate Society (ARES) in Tucson, USA (2025)
- the 31st Annual Conference of the European Real Estate Society (ERES) in Athens, Greece (2025)

Paper 3: Signals or Stories? ESG Sentiment and Investor Reactions in REIT Corporate Disclosure

Authors:

Sophia Bodensteiner, Lukas Lautenschläger, Prof. Dr. Wolfgang Schäfers

Submission Details:

Journal: Journal of Property Investment & Finance

Current Status: Revise and Resubmit (05/16/2026)

Conference Presentations:

This paper was presented at:

- the 41st Annual Conference of the American Real Estate Society (ARES) in Tucson, USA (2025)
- the 31st Annual Conference of the European Real Estate Society (ERES) in Athens, Greece (2025)
- the 32nd Annual Conference of the European Real Estate Society (ERES) in Vienna, Austria (2026)

Paper 4: News Sentiment and Stock Return Predictability: The Case of U.S. REITs

Authors:

Lukas Lautenschläger, Julian Lütticke, Prof. Dr. Wolfgang Schäfers

Submission Details:

Journal: Journal of Banking & Finance

Current Status: Submitted (04/15/2026)

Conference Presentations:

This paper was presented at:

- the 31st Annual Conference of the European Real Estate Society (ERES) in Athens, Greece (2025)
- the 32nd Annual Conference of the Pacific Rim Real Estate Society (PRRES) in Adelaide, Australia (2026)
- the 32nd Annual Conference of the European Real Estate Society (ERES) in Vienna, Austria (2026)

Awards

This paper was awarded the "Best PhD Paper Award 2026" at the 32nd Annual Conference of the European Real Estate Society (ERES) in Vienna, Austria (2026).

1.4 References

Baker, M., & Wurgler, J. (2006). Investor sentiment and the cross-section of stock returns. *The Journal of Finance*, *61*(4), 1645–1680. https://doi.org/10.1111/j.1540-6261.2006.00885.x

Baker, M., & Wurgler, J. (2007). Investor sentiment in the stock market. *Journal of Economic Perspectives*, *21*(2), 129–151. https://doi.org/10.1257/jep.21.2.129

Beracha, E., Lang, M., & Hausler, J. (2019). On the relationship between market sentiment and commercial real estate performance—A textual analysis examination. *Journal of Real Estate Research*, *41*(4), 605–638. https://doi.org/10.22300/0896-5803.41.4.605

Black, F. (1986). Noise. *The Journal of Finance*, *41*(3), 528–543. https://doi.org/10.1111/j.1540-6261.1986.tb04513.x

Bone-Winkel, S., Schäfers, W., Schulte, K.-W., Forster-Kraus, S., & Pfrang, D. (2008). Historische Entwicklung von REITs in globaler Perspektive. In S. Bone-Winkel, W. Schäfers, & K.-W. Schulte (Eds.), *Handbuch Real Estate Investment Trusts* (pp. 23–48). Immobilien Manager Verlag.

Braun, J., Hausler, J., & Schäfers, W. (2020). Artificial intelligence, news sentiment, and property market liquidity. *Journal of Property Investment & Finance*, *38*(4), 309–325. https://doi.org/10.1108/JPIF-08-2019-0100

Carstens, R., & Freybote, J. (2019). Tone in REIT financial statements and institutional investments. *Journal of Property Research*, *36*(3), 227–244. https://doi.org/10.1080/09599916.2019.1650802

Clayton, J., Ling, D. C., & Naranjo, A. (2009). Commercial real estate valuation: Fundamentals versus investor sentiment. *The Journal of Real Estate Finance and Economics*, *38*(1), 5–37. https://doi.org/10.1007/s11146-008-9130-6

De Long, J. B., Shleifer, A., Summers, L. H., & Waldmann, R. J. (1990). Noise trader risk in financial markets. *Journal of Political Economy*, *98*(4), 703–738. https://doi.org/10.1086/261703

Devlin, J., Chang, M.-W., Lee, K., & Toutanova, K. (2019). BERT: Pre-training of deep bidirectional transformers for language understanding. In J. Burstein, C. Doran, & T. Solorio (Eds.), *Proceedings of the 2019 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies, Volume 1 (Long and Short Papers)* (pp. 4171–4186). Association for Computational Linguistics. https://doi.org/10.18653/v1/N19-1423

Fama, E. F. (1970). Efficient capital markets: A review of theory and empirical work. *The Journal of Finance*, *25*(2), 383–417. https://doi.org/10.2307/2325486

Freybote, J. (2016). Real estate sentiment as information for REIT bond pricing. *Journal of Property Research*, *33*(1), 18–36. https://doi.org/10.1080/09599916.2016.1146791

Freybote, J., & Seagraves, P. A. (2017). Heterogeneous investor sentiment and institutional real estate investments. *Real Estate Economics*, *45*(1), 154–176. https://doi.org/10.1111/1540-6229.12132

Gallimore, P., & Gray, A. (2002). The role of investor sentiment in property investment decisions. *Journal of Property Research*, *19*(2), 111–120. https://doi.org/10.1080/09599910110110671

Grossman, S. J., & Stiglitz, J. E. (1980). On the impossibility of informationally efficient markets. *American Economic Review*, *70*(3), 393–408. https://www.jstor.org/stable/1805228

Hausler, J., Ruscheinsky, J., & Lang, M. (2018). News-based sentiment analysis in real estate: A machine learning approach. *Journal of Property Research*, *35*(4), 344–371. https://doi.org/10.1080/09599916.2018.1551923

Hirshleifer, D. (2001). Investor psychology and asset pricing. *The Journal of Finance*, *56*(4), 1533–1597. https://doi.org/10.1111/0022-1082.00379

Huang, A. H., Wang, H., & Yang, Y. (2023). FinBERT: A large language model for extracting information from financial text. *Contemporary Accounting Research*, *40*(2), 806–841. https://doi.org/10.1111/1911-3846.12832

Kahneman, D., & Tversky, A. (1979). Prospect theory: An analysis of decision under risk. *Econometrica*, *47*(2), 263–291. https://doi.org/10.2307/1914185

Keynes, J. M. (1936). *The general theory of employment, interest, and money*. Macmillan.

Koelbl, M. (2020). Is the MD&A of U.S. REITs informative? A textual sentiment study. *Journal of Property Investment & Finance*, *38*(3), 181–201. https://doi.org/10.1108/JPIF-12-2019-0149

Ling, D. C., Naranjo, A., & Scheick, B. (2014). Investor sentiment, limits to arbitrage and private market returns. *Real Estate Economics*, *42*(3), 531–577. https://doi.org/10.1111/1540-6229.12037

Lopez-Lira, A., & Tang, Y. (2023). *Can ChatGPT forecast stock price movements? Return predictability and large language models* [Working paper]. SSRN. https://doi.org/10.2139/ssrn.4412788

Loughran, T., & McDonald, B. (2011). When is a liability not a liability? Textual analysis, dictionaries, and 10-Ks. *The Journal of Finance*, *66*(1), 35–65. https://doi.org/10.1111/j.1540-6261.2010.01625.x

Marcato, G., & Nanda, A. (2016). Information content and forecasting ability of sentiment indicators: Case of real estate market. Journal of Real Estate Research, 38(2), 165–204. https://doi.org/10.1080/10835547.2016.12091442

Oh, J. S., & Verstein, A. (2024). A theory of the REIT. *The Yale Law Journal*, *133*(3), 755–838. https://yalelawjournal.org/article/a-theory-of-the-reit

Ruscheinsky, J. R., Lang, M., & Schäfers, W. (2018). Real estate media sentiment through textual analysis. *Journal of Property Investment & Finance*, *36*(5), 410–428. https://doi.org/10.1108/JPIF-07-2017-0050

Shiller, R. J. (2017). Narrative economics. *American Economic Review*, *107*(4), 967–1004. https://doi.org/10.1257/aer.107.4.967

Shleifer, A., & Summers, L. H. (1990). The noise trader approach to finance. *Journal of Economic Perspectives*, *4*(2), 19–33. https://doi.org/10.1257/jep.4.2.19

Shleifer, A., & Vishny, R. W. (1997). The limits of arbitrage. *The Journal of Finance*, *52*(1), 35–55. https://doi.org/10.1111/j.1540-6261.1997.tb03807.x

Tetlock, P. C. (2007). Giving content to investor sentiment: The role of media in the stock market. *The Journal of Finance*, *62*(3), 1139–1168. https://doi.org/10.1111/j.1540-6261.2007.01232.x

Walker, C. B. (2014). Housing booms and media coverage. *Applied Economics*, *46*(32), 3954–3967. https://doi.org/10.1080/00036846.2014.948675

2 Social Media and Real Estate: Do Twitter Users Predict REIT Performance?

2.1 Abstract

This study investigates the impact of social media sentiment on indirect real estate market returns by utilizing a comprehensive natural language processing approach to identify relevant Twitter posts and extract sentiment from them. To handle the complex linguistic features inherent in social media messages, three different sentiment classifiers are compared. The findings suggest a significant relationship between monthly sentiment and REIT returns, which occurs in two phases: a short-term speculative reaction and a greater longer-term reaction related to actual changes in the real estate market. The study also highlights that while the conventional dictionary approach can identify this relationship, more sophisticated classifiers can achieve higher accuracy. Overall, the results demonstrate the valuable insights that can be gained from analyzing social media data and its potential impact on the real estate market.

Keywords: Social Media Sentiment, Textual Analysis, Real Estate Investment Trusts, Machine Learning

2.2 Introduction

Social media has emerged as an increasingly important source of information in today's digital age. Since millions of users share their ideas, opinions, and experiences online, the platforms have become essential for businesses and individuals to gain information and communicate with their audience (Yadav & Vishwakarma, 2020). Thanks to widespread internet access, social media platforms have become incredibly convenient and accessible. As a result, users are immediately and always provided with multiple perspectives about any event worldwide. These technical advancements have fundamentally changed how people consume information. Even journalists increasingly rely on social media as a primary source for reporting (e.g., Broersma & Graham, 2013; Paulussen & Harder, 2014), highlighting its growing influence on today's society.

Consequently, social media provides insights into the opinions of those who actively share their thoughts online and reaches people who consume information directly on those platforms or indirectly via traditional sources. This wide reach of social media, in combination with a significantly growing exchange of stock market-related information on social media, has attracted the interest of researchers and business organizations in recent years. Sentiment analysis has found that social media's effect on stock markets is becoming more salient (Li et al., 2018) and utilized to predict stock market trends. In the finance literature, as early as 2011, Bollen et al. found evidence for public sentiment, obtained from social media, being a valuable predictor of stock market performance. Since then, many more researchers have found significant relationships between social media and stock markets (e.g., Sprenger et al., 2014; Xu & Cohen, 2018).

However, these findings have limited implications for real estate markets. Unlike the efficient stock markets, real estate markets are characterized by their heterogeneity, low divisibility, and high investment volumes, leading to segmented submarkets and long transaction periods. Therefore, the short-term social media analysis typically applied to stock markets has limited value for protracted real estate markets. Despite these limitations, there is little research on social media in the real estate literature and none which deals with it at a greater level, caused by the complexity of social media data. Two significant issues can describe this complexity: First, researchers must identify relevant tweets, which is difficult due to the lack of a consistent classification by topic of social media messages and the existence of non-user-generated content. Second, researchers must handle the linguistic complexity of social media messages, characterized by the use of informal language, shortened words and hashtags, and sarcasm and irony. These factors make it difficult to extract sentiment accurately, as traditional sentiment analysis techniques developed for news articles and corporate disclosures may not be suitable.

Therefore, this study provides a comprehensive approach to social media analysis, which includes an adoptable selection process and the comparison of three different sentiment classifiers.

To simplify the selection of tweets in the first place, this article investigates their relationship with Real Estate Investment Trusts (REITs), which can be traced using firm identifiers. As with previous social media sentiment literature, we expect social media can affect people's behavior and, therefore, shape economic outcomes. Consequently, we believe social media helps predict future REIT returns and that prediction accuracy can be increased using more sophisticated sentiment classifiers. In our study, we find that social media provides valuable insights into real estate market performance.

The results show that changes in social media sentiment cause a significant effect, which can be divided into two phases. The short-term reaction could be related to changes in the stock market, and the more substantial long-term effect is probably related to actual real estate market changes. Overall, the observed lag structure is similar to the relationship between sentiment expressed in the media and real estate markets, as Hausler et al. (2018) and Braun et al. (2020) observed. However, in line with our argumentation, the results indicate that social media sentiment is an even earlier indicator, which adds to our market understanding. This holds for all sentiment classifiers applied, even though we observe that more sophisticated approaches yield better classification performance and are better predictors.

In summary, similar to the findings of Li et al. (2018), we believe social media sentiment indicators can supplement traditional sentiment indicators. The accessibility, reach, and timeliness of the information on social media make it a powerful tool for predicting real estate market trends. Being able to quantify the sentiment expressed on social media, therefore, provides valuable insights to stakeholders in the real estate industry. Section 2.3 presents an overview of the sentiment literature focusing on real estate and social media. It also introduces the complexity of social media data and derives the hypotheses. Sections 2.4 and 2.5 show the data utilized and explain the methodology. Section 2.6 presents the empirical analysis results, and the conclusion in Section 2.7 summarizes the study's most significant findings, their implications for practice, and provides suggestions for future research.

2.3 Literature Review and Hypothesis Development

2.3.1 Sentiment in Real Estate

The impact of sentiment on real estate markets has been the subject of empirical studies for more than a decade, with a growing body of evidence suggesting a correlation between sentiment and market performance. As far back as 1936, Keynes observed that a lack of information leads market participants to supplement the limited information available with their

subjective perceptions, which he called the "animal spirit." Given that the real estate industry is frequently deemed inefficient owing to fragmented markets, property illiquidity, and a high information asymmetry (Gallimore & Gray, 2002), sentiment should have an influence on market performance. In 2009, Clayton et al. were among the first to find a positive relationship between sentiment and direct commercial real estate markets, and Lin et al. (2009) discovered similar results in indirect real estate markets of REITs. Both studies employed traditional survey-based (direct) sentiment indicators, which have since been adopted by other real estate researchers (e.g., Das et al., 2015a; Freybote, 2016; Ling et al., 2014). Those direct indicators extract sentiment through interviews and surveys, however, they face criticism due to being costly, time-delayed, and suffering from potential selection bias. Simultaneously, another line of research has emerged, utilizing (indirect) sentiment indicators derived from market fundamentals. These indicators operate on the premise that sentiment corresponds to what cannot be explained by market data (Baker & Wurgler, 2007) and have been applied multiple times in the literature (e.g., Freybote & Seagraves, 2017; Heinig et al., 2020; Marcato & Nanda, 2016). However, sentiment indicators derived from market fundamentals are referred to as indirect sentiment measures, as they depend on hard market data rather than soft sentiment components.

With both types of sentiment indicators having limitations, a new category of sentiment measure has become increasingly widespread. In 2007, Tetlock pioneered textual sentiment analysis in the finance literature, deriving a sentiment measure from qualitative information from news articles in The Wall Street Journal. He used a dictionary approach, identifying positive and negative words using predefined word lists, and determined the article's overall sentiment based on the ratio. As the results of this approach are highly dependent on the dictionaries used, the discovery by Loughran and McDonald (2011) of the relevance of a domain-specific dictionary marked another significant milestone in textual analysis. Subsequently, Walker (2014) was among the first researchers to utilize textual sentiment analysis in the context of real estate housing markets, examining news media articles concerning the UK housing market. Walker's finding that media Granger-causes house price changes has been confirmed by Soo (2015) using a simple frequency-based method on news media articles and by Ruscheinsky et al. (2018a), who created a domain-specific dictionary for the German market. Nowak and Smith (2017) took it a step further by incorporating text data from MLS property listings into a hedonic pricing model for house price valuation, which significantly enhanced the accuracy of their pricing model.

Textual analysis has been applied to housing markets and to commercial real estate markets. Beracha et al. (2019) demonstrated that sentiment within news abstracts of The Wall Street Journal can forecast the performance of the U.S. private commercial real estate market. Using a more advanced machine learning technique to measure sentiment, Hausler et al. (2018) revealed that sentiment extracted from professional financial news leads the performance of the U.S.

securitized and commercial real estate markets. Additionally, Plößl et al. (2023) improved prediction accuracy by categorizing news articles according to property type, thereby creating asset-class-specific sentiment measures. While previous studies focused on direct real estate markets, Ruscheinsky et al. (2018b) turned to indirect real estate markets. Using news media headlines, they found a significant leading relationship between sentiment and future REIT market movements. However, compared to direct real estate markets, indirect real estate markets with corporate disclosure offer an additional potential source for textual analysis. Therefore, Carstens and Freybote (2019) investigated the response of institutional REIT investors to abnormally positive sentiment in REIT financial statements. In a later study, Carstens and Freybote (2021) used those statements to predict commercial real estate total returns in the next quarter. Paulus et al. (2022) extended the research on textual sentiment by utilizing IPO filings to predict the underpricing of REIT IPOs. Alternatively, Price et al. (2017) investigated the vocal cues of CEOs during quarterly REIT conference calls to predict investors' reactions.

Similar to the concept of sentiment, Da et al. (2011) developed a direct measure of investor attention using aggregated search frequency in Google. This measure, using Google search frequency, is superior to other investor attention proxies derived from news coverage, since a search request is an active engagement, whereas a news article does not necessarily require investor interest nor is guaranteed to be read. Da et al. (2011) find that an increase in search volume predicts higher stock prices within two weeks. In a real estate context these findings are confirmed by Hohenstatt et al. (2011), who observe that on a national level real estate agency related search queries are a very robust indicator of transactions and that home financing related queries provide insights into the corresponding financing decisions. Similarly, at the metropolitan level, Das et al. (2015b) show that consumers' real estate searches are significantly associated with market fundamentals. These results hence show that there is a relationship between internet respectively social media data and real estate markets; however, those proxies do not consider the content of the articles or messages.

To summarize, previous studies indicate that real estate markets are influenced by human behavior and that textual analysis is a valuable tool for measuring sentiment. Nevertheless, despite analyzing different textual sources, such as newspapers and corporate disclosure, and promising evidence from investor attention analysis, there is a scarcity of research on the role of social media sentiment in real estate markets. There are only a few studies that look at its impact, all of them using direct housing real estate markets. Zamani and Schwartz (2017) collected 131 million tweets from U.S. counties based on self-reported locations to improve the prediction of foreclosure rates and price increases compared to traditional socioeconomic factors. Similarly, Hannum et al. (2019) explored the connection between microblogging sentiment and property prices in Istanbul by analyzing two million tweets marked with either specific geotags or spatial

lexical terms. They discovered a negative correlation between Twitter sentiment and property prices, including price appreciation. Recently, Tan and Guan (2021) investigated nine million tweets from Manhattan, New York City, and found that high sentiment levels aligned with high-frequency spatial clusters and higher housing prices. These studies highlight the predictive power of sentiment expressed in social media on local housing markets on a long-term perspective. However, they are limited in scope, as they focus solely on local markets and have yet to be applied to larger markets or explored broader market sentiment.

2.3.2 Social Media Sentiment in Finance

While the real estate literature has yet to explore the impact of social media extensively, there is a rapidly growing body of research in finance dedicated to understanding its influence. Social media sentiment analysis has emerged as a valuable source due to its real-time availability and increasing message volume related to stock markets. Studies have analyzed various online sources, including Twitter, StockTwits, and other message boards, with Antweiler and Frank (2004) being among the first to do so. Their study utilized a naïve Bayes classifier and a support vector machine (SVM) to analyze over 1.5 million message board postings on Yahoo! Finance and Raging Bull of 45 companies, revealing a link between disagreement in sentiment and increased trading volume. Additionally, the study found predictive power of their sentiment measure on next-day returns and volatility.

These early findings align with Bollen et al. (2011), who investigated whether collective mood states derived from large-scale Twitter feeds correlate with the value of the Dow Jones Industrial Average. They used two mood tracking tools to measure positive and negative mood, whereas mood is distinct into six dimensions. The results indicate that specific public mood dimensions can significantly improve the accuracy of their predictions. Chen et al. (2011) found evidence for a relationship between social media sentiment and stock markets and compared the effect with that of news media. Comparing The Wall Street Journal and Seeking Alpha, they state that there is a significant relationship in both cases, but social media has an even stronger and longer-lasting effect on the stock market.

Sprenger et al. (2014) diverged from previous research that concentrated on investigating the connection between social media sentiment and market indices, as they examined the correlation to individual stocks. Using a dictionary-based sentiment classifier, they found relationships between the bullishness in tweet sentiment and stock returns, message volume and trading volume, and disagreement and volatility. Especially, their findings about bullishness in social media have since been revisited. Li et al. (2018), for example, used a more sophisticated SVM and found that the bullishness in stock-related microblog messages had an impact on contemporaneous abnormal returns. Besides, they confirm that message volume significantly

correlates with trading volume and volatility. Overall, the literature suggests that there are significant short-term relationships between social media sentiment and stock markets.

In recent years, there has been a surge of research aimed at improving the handling of complex linguistic features in social media messages to enhance classification and prediction accuracy. This trend is attributable to the increased computing capacity now available, which enables the application of more sophisticated classifiers to larger datasets, ultimately resulting in better prediction accuracy. Therefore, Yildirim et al. (2019) compared deep learning and machine learning methods based on their classification performance of StockTwits messages and found that deep learning approaches outperform classic machine learning. These findings are supported by Shi et al. (2021) and Wang et al. (2020), which imply that different deep learning models outperformed other tested classical machine learning models. A recent study by Gupta and Chen (2020) examined the influence of financial-related social media messages on major companies' stock prices. They were able to improve prediction accuracy when including sentiment data. In conclusion, the literature shows that sentiment expressed on different social media platforms can be used to predict stock returns, volatility, and trading volume. Besides, the results suggest that deep learning models outperform classical machine learning models in analyzing social media sentiment.

2.3.3 Complexity of Social Media Data

Substantial evidence suggests the utility of social media sentiment in comprehending financial markets. However, the limited literature on real estate sentiment may be due to the multiple aspects of deriving sentiment from social media. Maynard et al. (2012) highlight significant challenges accompanying social media-based texts, which include issues of relevance, target identification, negations, contextual information, and volatility over time.

The issue of relevance arises when attempting to extract sentiments regarding a specific topic from social media texts. The corpus provided by social media platforms is enormous. However, not all of the messages are relevant for the investigation. Therefore, selecting the right messages is crucial since false selection will produce no or inaccurate results. Closely related to relevance is the issue of target identification. Ensuring that the sentiment is expressed about the analyzed topic can be difficult, especially when the topic is discussed in a broader context. This challenge requires the development of algorithms and tools that can identify the specific target topic of interest and exclude irrelevant discussions within the previously selected corpus. Besides those message selection issues, negations pose a significant challenge in textual analysis, particularly in dictionary-based sentiment analysis.

In addition, social media messages are characterized by opinions expressed through irony, sarcasm, and other semantic features, making it even more difficult for conventional sentiment

analysis to classify messages accurately. Hence, contextual information is required to understand those linguistic features and give a correct prediction for the expressed opinion. And not least, social media sentiment is exposed to strong temporal dynamics, resulting in volatility over time. Opinions expressed on a social media platform like Twitter can change radically from positive to negative and vice versa in a short period, which is why the time dimension is crucial for further analysis.

2.3.4 Hypotheses

Following Maynard et al. (2012), who emphasize the importance of relevance and target identification in social media analysis, we have chosen to focus on indirect real estate markets as they present a more manageable task of identifying relevant social media messages. Companies are often tagged using the company name or firm identifiers, which enables us to retrieve those messages. Besides, we chose to investigate the long-term relationship between social media sentiment and real estate markets rather than short-term relationships, due to the perception that real estate markets are inefficient and slow moving compared to stock markets.

This approach provides three additional benefits, such as comparing results to the abovementioned real estate studies, which typically examine monthly or quarterly relationships. Second, it aligns with Shiller's (2017, 2020) narrative economics, which suggests that beliefs and expectations about the future take time to form. Accordingly, the behavior of social media users or investors, even those who only monitor or indirectly consume social media via news media, could be influenced if the change in sentiment persists for an extended period. Third, investigating longer periods might reduce the high volatility issue in daily sentiment (Maynard et al., 2012). Despite that, we assume social media can be a valuable sentiment indicator since it more directly represents people's opinions compared to traditional news sources. Therefore, our first hypothesis is:

H1: *The sentiment expressed in social media affects future REIT market returns.*

Besides, analyzing indirect real estate with social media offers a unique lens through which one can discern the inefficiencies inherent in financial markets. In theory, financial assets are expected to reflect the performance of the real assets they represent. However, since real estate markets are slower to adjust to new information than financial markets, it is likely that a short-term fluctuation will occur that is not directly linked to the underlying real estate market. These short-term fluctuations may be caused by informed investors who have anticipated changing market conditions and are trying to offset the costs associated with being informed, as explained by Grossman and Stiglitz (1976). They further state that markets, in contrast to the efficient market hypothesis, do not fully reflect all available information immediately, as those are spread to uninformed investors passively through the information reflected within the price developments.

Hirshleifer and Teoh (2003), therefore, describe the idea of information cascading and herd behavior of investors in financial markets. These considerations can be seen as an extension of the idea of informed and uninformed investors by Grossman and Stiglitz (1976). The investor behavior described by Hirshleifer and Teoh (2003) does further support the idea of a delayed reaction that is induced by an earlier market movement. Therefore, the following is expected:

H2: *The sentiment expressed in social media is related to short-term and long-term fluctuations in REIT market returns.*

In light of Maynard et al.'s (2012) emphasis on the linguistic features inherent in social media messages and the aim to establish a baseline for future research, this study advocates for the application of more advanced techniques that can handle the challenges presented. Accordingly, the performance of different sentiment classification methods is evaluated. Consistent with prior literature, this study encompasses the full spectrum of classifiers, encompassing a dictionary-based, a machine learning, and a deep learning approach. Hence, the third hypothesis can be formulated as:

H3: *The more sophisticated a classifier, the more accurate its classification performance.*

As the accurate classification of tweets provides a foundation for the calculation of sentiment indices, it is essential to achieve a high level of accuracy in sentiment classification for reliable prediction of market performance. Since the sentiment indices are expected to reflect changes in the underlying market performance, a higher classification accuracy will result in more accurate predictions of market trends. Therefore, we hypothesize that higher classification accuracy will lead to better predictions of market performance. We, therefore, expect the following:

H4: *The more sophisticated a sentiment classifier, the better its predictions of future REIT market returns.*

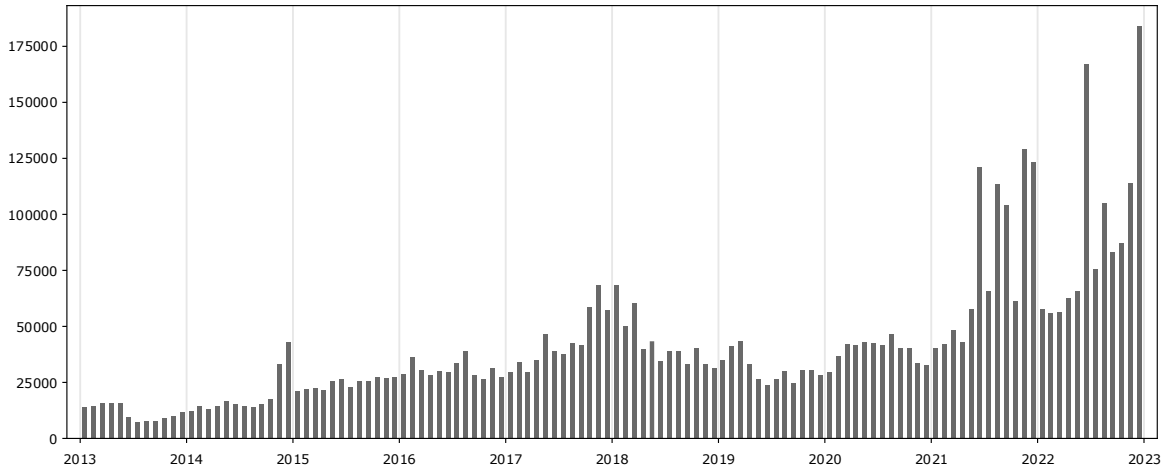
2.4 Data

2.4.1 Social Media Data

The corpus of this study consists of 4.97 million Twitter messages spanning a 10-year horizon ending in December 2022. Due to its popularity, we decided to use Twitter as our source, as it provides a large and diverse sample of data for sentiment analysis. Besides, Twitter has proven to be a valuable source of sentiment analysis in the finance literature, particularly regarding investor sentiment indicators. Therefore, the tweets were obtained from the Twitter Developer API using the name or cashtag (i.e., a "$" sign followed by the ticker symbol) of REITs. Given that we seek to compare sentiment indices with a representative performance index, the sample is restricted to tweets about REITs that have been listed in the performance index for the corresponding year. This approach ensures that the data used for comparison is both relevant

and comparable, allowing for a robust examination of the relationship between sentiment and performance.

Figure 2.1 – Unfiltered Tweet Distribution over Time



The monthly distribution of tweets, as depicted in Figure 2.1, reveals an overall increase in the number of messages from 2013 to 2020, followed by exponential growth until the end. This trend is not aligned with Twitter's monthly user base but rather with the onset of the global COVID-19 pandemic, indicating increased interest in REITs since then.

2.4.2 Dependent Variable

We used the U.S. REIT market to investigate the relationship between social media sentiment and indirect real estate markets. Being the largest REIT market in the world diminishes the impact of outliers on overall performance, which makes it more suitable for analysis. The FTSE Nareit All REITs Total Return Index indicates the market performance (Nareit, 2022). The National Association of Real Estate Investment Trusts (Nareit, www.reit.com) provides the index, which aggregates the total returns of all U.S. REITs. It represents all REIT types and includes price returns and income returns from dividends.

Descriptive statistics of the index can be found in Table 2.1.

Table 2.1 – Descriptive Statistics

	Mean	Median	SD	Min	Max
Total Return	0.0067	0.0112	0.0479	-0.2085	0.1142
Spread	0.0091	0.0089	0.0022	0.0055	0.0170
Term	0.0136	0.0141	0.0082	-0.0055	0.0296
Inflation	0.0248	0.0176	0.0227	-0.0020	0.0906

Note: Monthly total return as provided by Nareit. Macroeconomic control variables are the spread between Moody's Seasoned Baa- and Aaa-rated corporate bonds (*Spread*), the difference between the 10-year Treasury bond yield and the 3-month Treasury bill (*Term*), and the inflation rate of the U.S. (*Inflation*).

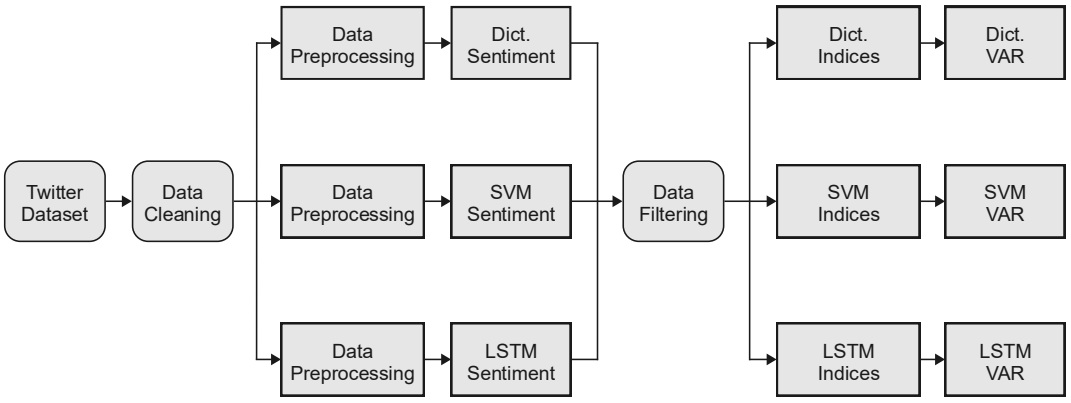
2.4.3 Economic Control Variables

When performing regression analysis for real estate returns, it is crucial to consider other factors that may impact, such as the overall economy and capital markets (Brooks & Tsolacos, 2010). In academic literature, various exogenous variables have been employed to explicate the performance of indirect real estate. For this study, we use the most commonly utilized variables. As a metric to represent the overall economic default risk, this study includes the spread between Moody's Seasoned Baa- and Aaa-rated corporate bonds (*Spread*), as established in the literature by Lin et al. (2009), Ling et al. (2014), Freybote and Seagraves (2017), and Hausler et al. (2018). To account for expectations regarding future economic development, a macroeconomic control in the form of the difference between the 10-year Treasury bond yield and the 3-month Treasury bill (*Term*) is included, as previously utilized by Clayton et al. (2009), Freybote and Seagraves (2017), and Hausler et al. (2018). Finally, following Hausler et al. (2018), the regression accounts for the inflation rate of the U.S. (*Inflation*) as real estate is commonly regarded as a possible hedge against the change in the consumer price index (Hoesli et al., 2008). Descriptive statistics of the variables can be obtained from Table 2.1.

2.5 Methodology

Examining the relationship between Twitter messages and REIT market performance is a process that involves various techniques of natural language processing. This complex process is illustrated in Figure 2.2 and can be divided into three phases. The first phase involves the selection of tweets and begins with the retrieval of messages from Twitter. Those messages are then filtered to obtain a corpus of relevant tweets. During this phase, the textual corpus undergoes preprocessing. In the second phase, the filtered tweets are assigned to one of three sentiment categories (positive, negative, or neutral) by applying three classification techniques. Once all of the messages have been classified, monthly and quarterly sentiment indices are generated and, in the final phase, integrated into econometric models to evaluate their relationship with REIT market performance.

Figure 2.2 – NLP Process



2.5.1 Tweet Selection and Data Cleaning

In light of Maynard et al.'s (2012) research on the importance of relevance and target identification in social media sentiment analysis, this study places substantial emphasis on the selection of tweets about all U.S. REITs. To this end, a three-step procedure is employed for tweet selection: First, English tweets containing the name or cashtag (i.e., a "$" sign followed by the ticker symbol) of a REIT are downloaded. Second, tweets without interactions, such as likes, comments, or retweets, as well as duplicates, are removed from the dataset (compare Sprenger et al., 2014). Third, drawing on the methodology proposed by Zhou et al. (2017), a word embedding approach is utilized to categorize tweets and eliminate irrelevant or non-pertinent topics.

Upon completing the first step and examining the data, it is evident that a substantial proportion of the tweets downloaded are irrelevant to the research question (e.g., tweets in other languages or on unrelated topics) and that further selection is required. Tweets that only include the ticker symbol or appear to be generated by automated systems and typically do not have any interactions were removed from the dataset, as were all duplicates. This decision is in line with the argumentation about relevance put forth by Maynard et al. (2012) that tweets without interactions are likely to be irrelevant to the target audience, and the findings of Sprenger et al. (2014), demonstrating that users providing above average investment advice are retweeted more often, amplifying their share of voice. A closer examination of the distribution of tweets among REITs then reveals that some REITs appear to have higher levels of public interest at first glance. Upon further screening of these tweets, it becomes evident that some of them share ticker symbols with other financial products, such as cryptocurrencies, highlighting the importance of further data cleaning and the third step of our procedure.

Aiming to identify those tweets that are really about REITs, we, similar to Zhou et al. (2017), employ a combination of two techniques for word embedding and similarity measurement, namely the Global Vector (GloVe) model and cosine similarity. These algorithms allow tweets to be categorized by topic through signal words and then create a sentiment indicator only for "REIT" related tweets. This automated procedure aligns with the findings of Watanabe and Zhou (2020), who determined that signal words should possess both knowledge-based and frequency-based properties. The GloVe model, introduced by Pennington et al. (2014), is a word embedding technique that generates a vector representation for each word in the corpus, capturing the semantic and syntactic properties of the words. To generate these, the algorithm combines the advantages of global matrix factorization and local context window methods.[1] The training of

[1] Matrix factorization techniques are used to decompose large matrices that contain statistical information about a text corpus, resulting in low-dimensional word representations and reduced computation time. On the other hand,

the algorithm is carried out on only non-zero elements in a global word-word co-occurrence matrix *X*, rather than on the entire sparse vector space or the local context windows within a large text corpus, and therefore generates a matrix of meaningful substructures. The objective function J of the weighted least squares regression model is specified as follows:

$$J = \sum_{i,j=1}^{V} f(X_{ij})(w_i^T \widetilde{w}_j + b_i + \tilde{b}_j - \log X_{ij})^2 \tag{2.1}$$

where V is the size of vocabulary; X_{ji} lists the number of times word j occurs in the context of word i; w ($\widetilde{w}$) represents the word vector for a main (context) word and b ($\tilde{b}$) describes the scalar biases for main (context) words; w^T indicates a transposed word vector. In order to prevent the learning of simple common word pairs, a weighting function readjusts the cost for each word pair because word pairs exhibit different occurrence frequencies in the text corpus. This weight is either set to 1 if the co-occurrence count exceeds a certain threshold or based on the co-occurrence count (see Pennington et al. (2014) for additional information).

The matrix obtained from the GloVe model consists of word vectors that identify relationships between words in the corpus. To measure the similarity between these vectors, we apply the cosine similarity, a nearest neighbor approach, which calculates the cosine of the angle between two vectors.

This method identifies synonyms or words used in similar contexts. Cosine similarity is computed as follows:

$$Cosine\ Similarity\ (v_i, v_j) = \frac{v_i \cdot v_j}{\|v_i\|\ \|v_j\|} \tag{2.2}$$

where $v_i \cdot v_j$ is the dot product of two GloVe vectors, $\| v_i\|$ is the length of vector v_i, and $\|v_j\|$ is the length of vector v_j. Cosine similarity is geometrically the cosine of the angle between the two vectors, normalized by the vector length. Since word vectors are non-negative, the measure is bounded between 0 and 1. Besides, a higher value indicates that the words are more semantically similar and tend to occur in the same context.

To create a meaningful dataset of tweets, we filter for tweets that include at least one of the 25 words best describing "REIT".[2] Hence, v_i represents the global vector of "REIT" and v_j is iterated through for all other words, with the 25 words yielding the highest cosine similarity selected as the signal words. Using signal words to filter tweets, instead of applying an entire topic model, as performed by Zhou et al. (2017), is a more efficient and targeted approach. Using signal words

context window methods use linear relationships between word vectors based on local context windows to predict linguistic patterns and improve word analogy tasks.

[2] It was determined that filtering for the 25 closest words was optimal as a significant decline in the quality of the word representations occurred beyond this threshold.

allows for a focused identification of relevant tweets, instead of the broad topic identification provided by a topic model such as the Latent Dirichlet Allocation (LDA) model. Furthermore, using signal words can reduce the computational complexity associated with topic models, making it a more practical solution for large-scale text analysis tasks. Nonetheless, future studies may seek to apply topic modeling for this purpose.

It is essential to mention that preprocessing of tweets was carried out between steps two and three to enhance the accuracy and efficiency of the classification and the upcoming sentiment analysis. This involved lowercasing, removal of punctuation, numbers, non-alphabetic and special characters, and stop words. Contractions were expanded, tweets were tokenized, and words ultimately lemmatized. Contractions can cause noise in the analysis since word frequencies are distorted, but their expansion resolves this issue (Dorle & Pise, 2018). Tokenization involves dividing the text into meaningful units such as phrases, n-grams, or, in this case, words. Ultimately, lemmatization (Manning et al., 2008), the conversion of words to their root form, was preferred over stemming as it is a more advanced method.

Besides, the literature suggests additional preprocessing for social media data (e.g., Ankit & Saleena, 2018; Bahrainian & Dengel, 2013; Dorle & Pise, 2018), of which the most commonly used steps were incorporated. These included the removal of @ mentions, the # symbol, URLs, retweets, non-ASCII letters, and symbols. The slang commonly used on social media platforms was addressed by replacing slang terms with their actual phrasal equivalents (Bahrainian & Dengel, 2013). This conversion reduced the vocabulary size, thereby improving performance for dictionary-based approaches, as actual words are more likely to be in the dictionary than slang terms. The Internet & Text Slang Dictionary from www.noslang.com, which contains more than 5000 slang terms and their formal meaning, was provided for converting the slang (Jones, 2022).

2.5.2 Sentiment Analysis

After conducting preprocessing and selecting tweets, sentiment analysis is performed to determine the sentiment for each of the remaining tweets. This involves classifying tweets as positive, negative, or neutral and then aggregating those individual scores within specified time frames to derive a sentiment index.

Classification with a Dictionary Approach

In this study, the dictionary approach is employed as the first algorithm and serves as the baseline for evaluating the performance of subsequent methods. The dictionary approach was selected due to its widespread use as the most fundamental and commonly used method for sentiment analysis (e.g., Beracha et al., 2019; Carstens & Freybote, 2019; Paulus et al., 2022). It is a simple yet effective method brought into finance by Tetlock (2007). A predefined list of words, referred to as the sentiment dictionary, is used to assign sentiment polarity to text. Since the outcome of

sentiment analysis heavily depends on the dictionary used and, given that domain-specific idiosyncratic word meanings exist, the use of domain-specific dictionaries is recommended. Doran et al. (2012) and Henry and Leone (2016) have shown that domain-specific dictionaries help to account for these nuances and enhance the accuracy of sentiment analysis. Therefore, we apply the real estate-specific dictionary developed by Ruscheinsky et al. (2018b), which is based on the finance-related dictionary by Loughran and McDonald (2011) and has been adjusted for a real estate context.

Furthermore, the dictionary-based approach is also known as a bag-of-words approach because it disregards sentence structure. This simplicity does not make it an adequate baseline for comparison with more advanced algorithms, as it is the only one not capturing critical contextual information in sentiment analysis of social media content, as Maynard et al. (2012) described. To address this limitation, valence shifters, such as negators (e.g., not), amplifiers (e.g., very), de-amplifiers (e.g., less), or adversative conjunctions (e.g., but), are controlled for. In doing so, a context cluster is defined with n words before and after each polarized word. Within this context cluster, valence shifters are identified using an additional dictionary containing 76 valence shifters. Once a valence shifter has been identified, it affects the score of the polarized context cluster by a certain weight (see Rinker, 2019, for a detailed description).

Classification with a Support Vector Machine Model

Since machine learning approaches tend to yield higher classification accuracy than dictionary-based approaches (Li, 2010), we include an SVM approach. The SVM approach was selected due to its effectiveness in handling high-dimensional data and ability to learn complex decision boundaries. Besides, in real estate literature, it is among the commonly used machine learning approaches for textual sentiment classification (e.g., Hausler et al., 2018; Antweiler & Frank, 2004). The SVM, as developed by Cortes and Vapnik (1995), involves training a model using a labeled dataset[3] to predict sentiment polarity. Therefore, it is trained to map tweets, in the form of input vectors, into a high- dimensional space via a nonlinear mapping technique chosen a priori. This vector space is subdivided by a hyperplane, which subsequently allows identifying the category of tweets depending on their position in the feature space relative to the surface. However, since SVMs are binary classifiers and cannot classify into three categories, Platt scaling (Platt, 1999) is applied to get a probabilistic output, with *P* interpreted as the probability of a text being positive, we classify the tweets in three sentiments:

[3] The training dataset utilized in this study is sentiment140 and is classified by Go et al. (2009) using distant supervision. This balanced dataset contains 1.6 million tweets labeled as either positive or negative.

$$P \geq 0.667 \rightarrow \text{positive sentiment}$$
$$0.333 < P < 0.667 \rightarrow \text{neutral sentiment}$$
$$P \leq 0.333 \rightarrow \text{negative sentiment}$$

A detailed description of the SVM model can be found in Appendix I.

Classification with a Long Short-Term Memory Model

The third classification model chosen for this study is a long short-term memory (LSTM) model. LSTMs are modified versions of recurrent neural networks (RNNs) introduced by Hochreiter and Schmidhuber (1997) and are deep learning classifiers. They are well suited for sentiment analysis due to their ability to model the contextual dependencies within textual data effectively. The internal memory state of the model allows it to capture both short- and long-term dependencies in the text, improving its ability to accurately predict sentiment. LSTMs have been widely applied in sentiment analysis, demonstrating superior performance compared to traditional models (e.g., Fu et al., 2018; Jin et al., 2020; Wu et al., 2022). With ever more (social media) data available for training[4], they provide better scalability, which further deems them superior to traditional indicators.

A neural network is a multi-layer computational model composed of nodes organized into multiple layers. It comprises an input layer, several hidden layers, and an output layer (Sarker, 2021). The input is received as a document feature matrix at the input layer, then transformed through the application of weights and biases, and then passed to the first hidden layer. In each hidden layer, the transformed values are further transformed by an activation function. This process is repeated based on the number of hidden layers and nodes until the final transformed values are sent to the output layer. Ultimately, the values from the final hidden layer are aggregated to produce the final output.[5] In RNNs, the output values of a cell can also be added back to the inputs of the current layer. This is a major feature, called a feedback loop, that distinguishes them from other neuronal networks.[6]

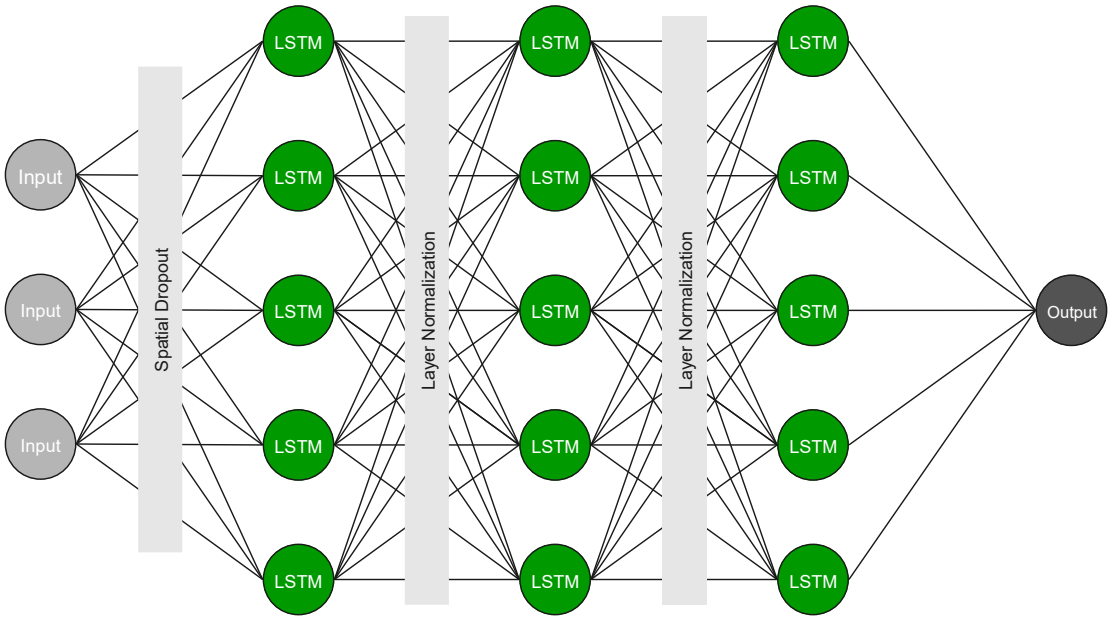
The LSTM differentiates itself from standard RNNs by the memory cell added in each LSTM node of the network. This memory cell evolves around its central feature, called the constant error carousel (CEC). The CEC is a mechanism where the cell state, the current value saved in the memory cell, is multiplied by a weight of 1 to keep the value constant. Additionally, an input and

[4] The same training data set and preprocessing steps as for the SVM have been applied.
[5] Since the output is again probabilistic, with P indicating the probability of an input having positive sentiment, the same classification ranges as for the SVM have been applied.
[6] Due to the feedback loop, RNNs encounter the challenge of vanishing or exploding gradients. As the weights are applied repeatedly to the values in the feedback loop, the magnitude of the input values either decreases (weights < 1) or increases (weights > 1) with every iteration, potentially leading to the input values either dropping to zero or increasing to infinity after a sufficient number of iterations. Vanishing or exploding gradients, however, impede the ability of the network to converge toward optimal weights and biases (Hochreiter, 1991).

an output gate are added. These gates are used to induce changes in or regulate the output of the internal cell state (Hochreiter & Schmidhuber, 1997). In modern implementations of LSTM networks, an additional gate unit called the forget gate is added. This gate acts as a way to reset the internal state as given by Gers et al. (2000). The usage of these three gates enables the LSTM cell to either save past weights and thus past information or alter these weights to overwrite the information saved. In general, this additional memory cell enables the ability of the LSTM to keep long-term information as the cell state can be kept constant. Additional information on the LSTM model applied in this study is presented in Appendix II.

The LSTM network used in this study is illustrated in a simplified version in Figure 2.3. The model consists of an input layer, three hidden LSTM layers with 256 nodes each, and an output layer. Additionally, layer normalization is included between each LSTM layer, improving the efficiency of a model (Ba et al., 2016).

Figure 2.3 – LSTM Model



Note: The LSTM model consists of an input layer, three hidden LSTM layers with 256 nodes each, and an output layer. Between each layer, layer normalization is included.

Model Accuracy

In this study, model accuracy is defined as the proportion of tweets that are correctly classified by our models, compared to the total number of tweets in the test set. To facilitate this, the sentiment140 dataset by Go et al. (2009), comprising 1.6 million pre- classified tweets (either positive or negative), is split randomly into two parts: 80% for training and 20% for testing.[7] This 20% segment is not used in the training process and is reserved exclusively for evaluating the classification performance of the models. We employ five independent models for both machine

[7] For the SVM, only 25% of the full training dataset is used for training and testing, due to computational limitations.

learning approaches to enhance the reliability of our findings. This methodology, involving unique training and test splits for each model, improves the robustness and generalizability of the classification results. The final sentiment label of a tweet is calculated as the average classification across these five models.

After implementing the three sentiment classifiers, their accuracy can be compared by assessing the proportion of correctly classified tweets from the separated 20% of the dataset (the test set). It should be noted that due to the lack of neutral classification in the training dataset, the accuracy of the classifiers is measured only for a binary classification. In the case of the dictionary approach, this implies that only tweets containing at least one polarized word are considered. For the other two approaches, a text is classified as negative if the probability of this text being positive (*P*) is below 0.5 and as positive if it is above 0.5. The accuracy is then determined by comparing the amount of correctly classified tweets to the overall number of tweets. For the dictionary approach, tweets classified as neutral are fully excluded.

Index Construction

After tweets are classified by all three models, the positive, negative, and neutral tweets are aggregated into sentiment indices. As we seek to understand the long-term relationship and face temporal dynamics in sentiment, we decided to consider different time horizons in the analysis. As such, we have chosen to employ monthly indices for our main analysis and quarterly indices to check the robustness of the results. The first measure, following Antweiler and Frank (2004), is called the Optimism Indicator (*OI*) and shows the level of optimism expressed on social media:

$$OI_t = \frac{\sum_{i=1}^{I} \text{positive message}_{i,t}}{\text{total number of messages}_t} \tag{2.3}$$

The second measure is called the Pessimism Indicator (*PI*). It is based on Rozin and Royzman (2001), who found that individuals are affected more strongly by negative rather than positive influences and therefore measures the ratio of negative messages:

$$PI_t = \frac{\sum_{j=1}^{J} \text{negative message}_{j,t}}{\text{total number of messages}_t} \tag{2.4}$$

The third measure is the Sentiment Quotient (*SQ*), which is the most used in literature and provides an indicator of the number of positive messages compared to positive and negative messages:

$$SQ_t = \frac{\sum_{i=1}^{I} \text{positive message}_{i,t}}{\sum_{i=1}^{I} \text{positive message}_{i,t} + \sum_{j=1}^{J} \text{negative message}_{j,t}} \tag{2.5}$$

with i as a positive message, j as a negative message, and t states the aggregated time frame.

2.5.3 Econometric Approaches

Vector Autoregression

A vector autoregression (VAR) framework is applied to examine the intertemporal relationship between social media sentiment regarding REITs and the performance of the corresponding REIT market. The VAR model offers an advantage in that it does not make a priori assumptions about causal relationships and allows for estimating the intertemporal relationship between sentiment and performance while controlling for possible endogeneity. The model specifies each variable as a linear function of its own lags and the lags of other variables, enabling control for momentum in the dependent variable (Beracha & Downs, 2015). However, following economic theory, exogenous variables should be incorporated into the VAR framework. In matrix notation, where X denotes a matrix of exogenous variables and B a matrix of coefficients, the model can be written as follows:

$$y_t = A_0 + \sum_{i=1}^{l} A_i \, y_{t-i} + BX_t + u_t \tag{2.6}$$

In the VAR framework, the stationarity of its components is tested using the Augmented Dickey-Fuller (ADF) Test. Whenever the null hypothesis of unit root presence is rejected, it indicates the need for differencing the variables to ensure statistical validity. Additionally, a suite of diagnostic tests is conducted to confirm the robustness of the results. The Breusch-Godfrey Lagrange Multiplier test is employed to verify that the residuals are not serially correlated. Furthermore, normality and heteroscedasticity tests are performed to guarantee the statistical validity of the results. Newey and West (1987) standard errors, which are robust to heteroscedasticity and autocorrelation (HAC), are applied whenever necessary.

Determining the optimal lag length is a crucial step in configuring the VAR. The optimal lag length is optimized using the dictionary approach and applied to the other models to ensure comparability among the various models. This is selected using the Akaike Information Criterion (AIC) as selection criterion and choosing the lag length that minimizes its value (Brooks & Tsolacos, 2010). The optimal lag length for the dictionary approach with three sentiment indicators (*OI*, *PI*, *SQ*) is determined by calculating the AIC for each indicator and ranking their values. The lag with the lowest mean rank is ultimately selected. Hereby, the minimum lag length is restricted according to recent findings. As Hausler et al. (2018) and Braun et al. (2020) found a significant impact of sentiment on real estate markets of up to eight months, this will serve as our minimum lag length in the monthly analysis.

Granger Causality and Impulse Response Function

To gain a deeper understanding of the causal relationship and the impact of social media sentiment on market performance, Granger causality will be assessed, and impulse response functions constructed.

Granger causality is a concept used in time series analysis to determine the causal relationship between two variables (Granger, 1969). The idea behind Granger causality is that if a variable *X* is said to Granger-cause a variable *Y*, it means that the past values of *X* provide information that helps predict future *Y* values beyond what can be predicted from the past values of *Y* alone. In other words, *X* is a useful predictor of *Y*. Thus, the use of Granger causality aids in comprehending the temporal relationships between social media sentiment and market performance, ultimately determining if sentiment holds predictive power and vice versa.

Impulse response functions are used to measure the impact of a sudden, short- lived shock to a particular variable (i.e., sentiment) on other variables (i.e., performance) within the system. Therefore, an exogenous event in the form of a one standard deviation shock in the sentiment variable is simulated, and its effect, i.e., the dynamic response of the total return, during the analyzed lags, is graphically represented. This provides further insights into the transmission mechanisms of the interdependent variables and the speed at which the shock decays over time.

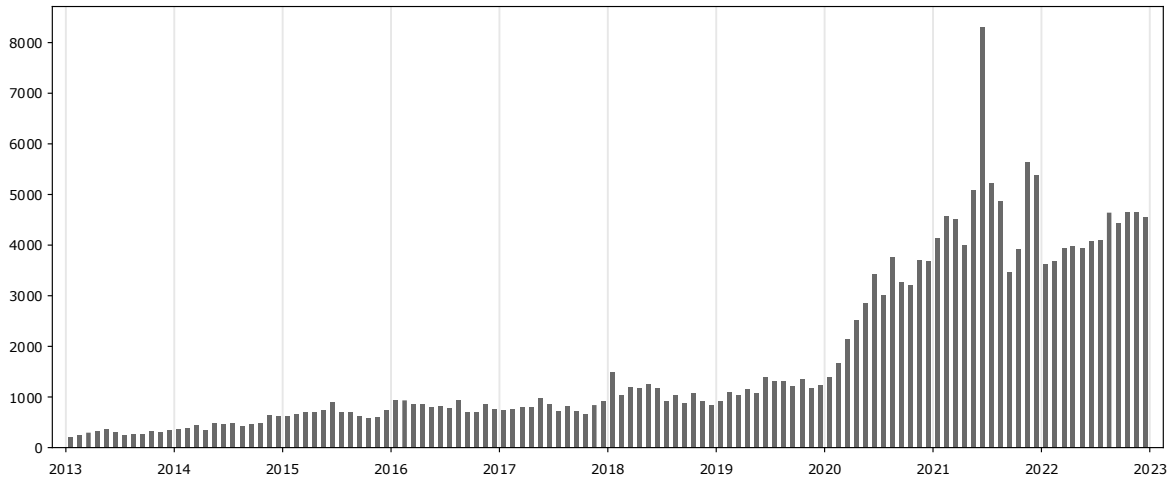
2.6 Results

2.6.1 Tweet Selection

Initially, the corpus of tweets consisted of more than 4.97 million tweets, of which a significant proportion was deemed irrelevant to the study. After removing duplicates and tweets with no interaction, the dataset was reduced to approximately 650,000 tweets and used to train the GloVe model. Finally, the corpus was pruned to 210,559 tweets, utilizing the nearest neighbors of "REIT" as signal words to ensure that the remaining messages exclusively pertained to REITs.[8] Figure 2.4 shows the distribution of the final selection during the observation period, which is similar to the distribution of all tweets as shown in Figure 2.1. The number of tweets generally increased until 2020, with a significant surge thereafter. This suggests that our selection process is not biased and selects a constant percentage of available tweets over time, which indicates this approach is generalizable.

[8] The list of signal words can be found in Appendix III.

Figure 2.4 – Filtered Tweet Distribution over Time



2.6.2 Model Accuracy

After tweet selection was accomplished and the sentiment classifiers trained, the accuracy of the three models is determined using the separated 20% of the pre-classified training dataset. According to our findings, the dictionary approach yielded the lowest accuracy of 75.64% in correctly classifying tweets. For the SVM, an accuracy of 79.56% and for the LSTM, an accuracy of 82.79% was reported.[9] Note that these accuracies refer to the labeled training dataset, and not the actual tweet dataset used in this study. Nevertheless, they give a good impression of the capabilities of the different methods and suggest that more sophisticated models generally lead to better classification performance, supporting the hypothesis that more sophisticated classifiers are more accurate. To further improve the performance, it is important to note that incorporating a neutral category in our actual model mitigates the impact of wrongly classified tweets and tweets that are not polarized.[10]

2.6.3 Sentiment Indices

Figures 2.5 to 2.7 show the *OI*s, *PI*s, *SQ*s, and the total return index during the observation period. Upon visual inspection, it is evident that the dictionary approach is the least optimistic, which may be attributed to the limited number of positive words in the dictionary. Besides, the three classifiers exhibit a relatively low standard deviation, which, for example, can be seen at the beginning of the COVID-19 pandemic in early 2020, where only a marginal and short-lasting slump in optimism occurs. Moreover, the standard deviation of the *OI*s diminishes over time,

[9] The model accuracy of the SVM and LSTM increased from 76.22%, respectively, 78.02% when stop words were excluded, which confirms our assumptions described in the data-cleaning section. Further note that the accuracy of the LSTM is given by the validation process after the final step of the training process in contrast to a separate testing as conducted for the SVM.

[10] Since it is impossible to measure the accuracy of our unlabeled corpus, we compared the percentage of tweets that were classified identically by the models. The LSTM model and dictionary shared 40.44% similarity, the dictionary and SVM shared 48.80%, and LSTM and SVM shared 61.03%. For the dictionary approach, significantly more tweets have been classified neutral due to the limited size of the word lists.

which should be related to the increasing number of tweets available. Looking at the development over the entire observation period, it seems there is a correlation between optimism and total returns.

Figure 2.5 – Optimism Indicators and Monthly Total Return over Time

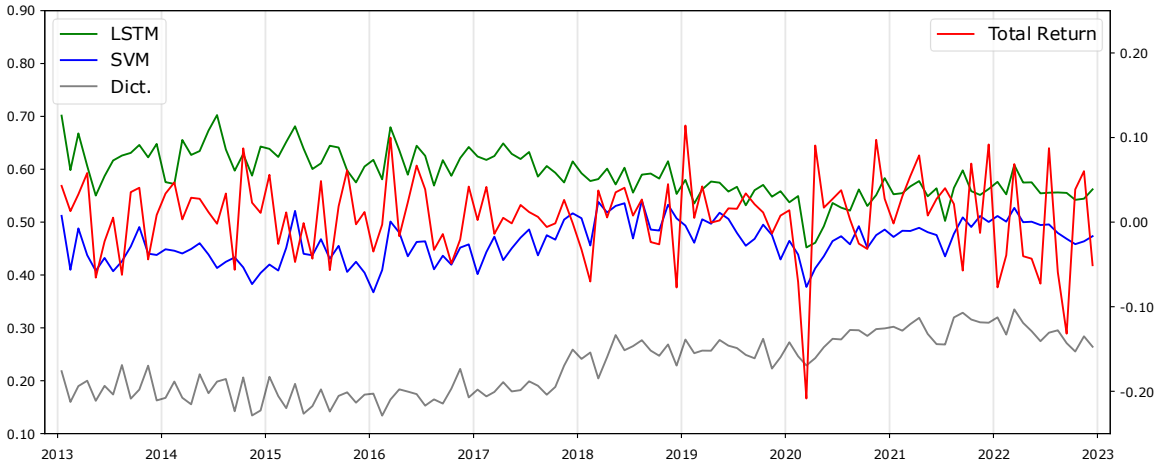


Figure 2.6 – Pessimism Indicators and Monthly Total Return over Time

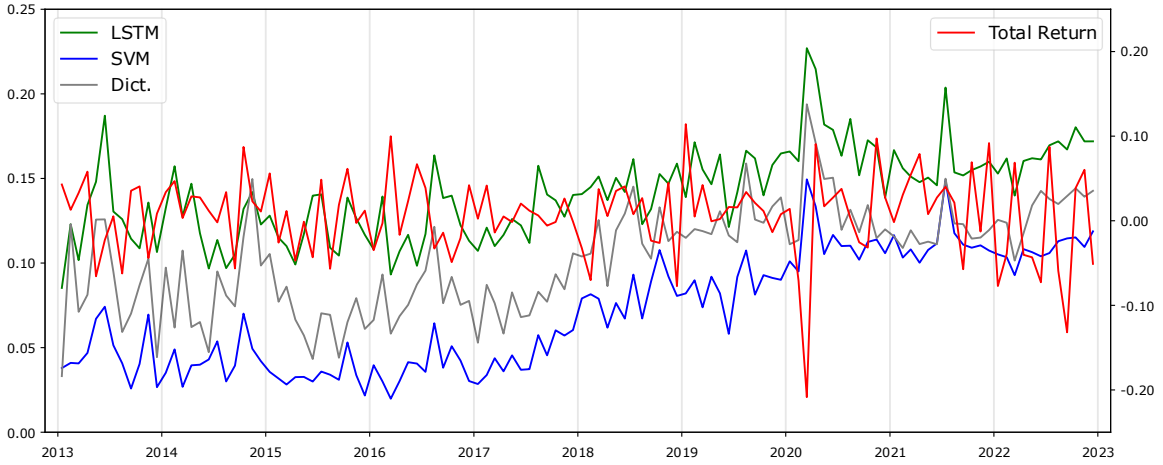
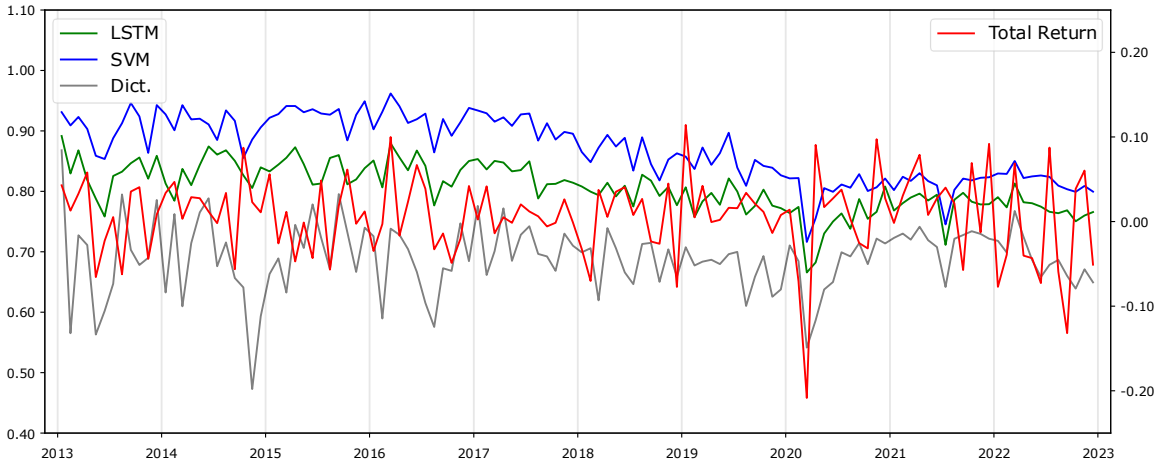


Figure 2.7 – Sentiment Quotients and Monthly Total Return over Time



Regarding the *PI*, as indicated by Figure 2.6, the indices are closer together but have a higher standard deviation. For all three models, there was a clear increase in pessimism at the start of the pandemic, but similar to the *OI*, it quickly returned to normal levels. Moreover, the models have identified a rise in pessimism starting from 2017, which has stagnated since the pandemic. This lack of change in recent years is understandable, given the increased volatility of total returns and the challenging economic conditions. The *SQ* is, by definition, connected to the *OI* and the *PI*. Therefore, the previously mentioned findings are reflected in it. Hence, the visual analysis supports our first hypothesis of social media sentiment affecting REIT market returns.

2.6.4 Vector Autoregression

Table 2.2 presents the results of the regressions, including the *OI* as the endogenous sentiment variable, for the three classifiers. Both the dictionary approach and the LSTM model exhibit a short-term inverse relationship between optimism and total return, which seems counterintuitive at first. An increase in optimism tends to result in a decrease in total return in the following month. Interestingly, this short-term inverse relationship is similar to the relationship observed between total return today and total return in the previous month.

Table 2.2 – VAR Results: Monthly Lags Optimism Indicator

	FTSE Nareit All REITs Total Return		
	OI Dict.	**OI SVM**	**OI LSTM**
Total Return$_{t-1}$	-0.253**	-0.274**	-0.252*
Total Return$_{t-2}$	-0.068	-0.046	-0.052
Total Return$_{t-3}$	0.037	0.015	0.058
Total Return$_{t-4}$	0.012	0.068	-0.013
Total Return$_{t-5}$	0.114	0.117	0.096
Total Return$_{t-6}$	-0.097	-0.133	-0.057
Total Return$_{t-7}$	0.029	0.037	-0.027
Total Return$_{t-8}$	0.065	0.129	0.069
OI$_{t-1}$	-0.362*	0.019	-0.261**
OI$_{t-2}$	-0.177	-0.061	0.039
OI$_{t-3}$	0.058	0.172	0.132
OI$_{t-4}$	0.171	-0.014	0.100
OI$_{t-5}$	-0.042	-0.056	-0.057
OI$_{t-6}$	0.404	0.198	0.205
OI$_{t-7}$	0.289	0.164	0.400***
OI$_{t-8}$	0.281	0.022	0.112
Spread	-17.001***	-16.472***	-17.068***
Term	-5.970**	-7.215***	-7.141***
Inflation	1.145	0.996	0.958
Constant	0.006	0.005	0.007
Adj. R²	0.186	0.149	0.219
	Granger Causality		
Sentiment	0.288	0.714	0.088
Total Return	0.667	0.524	0.125

Note: This table shows the estimated coefficients from the VAR models with monthly total returns and the Optimism Indicator. The set of the macroeconomic control variables includes the spread between Moody's Seasoned Baa- and Aaa-rated corporate bonds (*Spread*), the difference between the 10-year Treasury bond yield and the 3-month Treasury bill (*Term*), and the inflation rate of the U.S. (*Inflation*). All variables, except for total return, are considered as the first differences. Newey and West (1987) standard errors, which are robust to heteroscedasticity and autocorrelation, are used. Granger causality p-values are reported for both directions. * denotes significance at the 10% level, ** at the 5% level, and *** at the 1% level.

This could be interpreted as investors becoming skeptical about the sustainability of the returns and, therefore, anticipating a downward correction. The high variance observed in the earlier lags also might indicate an imbalance in the level of information between different investors, supporting Grossman and Stiglitz (1976). Despite this short-term reversal effect, the LSTM model detects a long-term relationship significant at 1% level between optimism and total returns in lag seven. The coefficient has the expected sign, and its economic significance is the highest among the significant coefficients, therefore, having the biggest impact on returns.

As Table 2.3 indicates, the results for the *PI*s are even more promising. Again, the dictionary approach identifies an inverse relationship in the first lag between pessimism and total returns, emphasizing the previously described skepticism among investors and an expectation of a downward correction. The inverse relationship is also consistent with the total return in the previous month. However, in contrast to the *OI*, the LSTM model finds a short-term relationship between pessimism and total returns in lag two and three, which has the expected sign.

Table 2.3 – VAR Results: Monthly Lags Pessimism Indicator

	FTSE Nareit All REITs Total Return		
	PI Dict.	**PI SVM**	**PI LSTM**
Total Return $_{t-1}$	-0.285**	-0.244*	-0.299**
Total Return $_{t-2}$	-0.068	-0.069	-0.092
Total Return $_{t-3}$	0.063	0.093	0.044
Total Return $_{t-4}$	0.057	0.070	-0.006
Total Return $_{t-5}$	0.110	0.062	0.040
Total Return$_{t-6}$	-0.111*	-0.088	-0.091
Total Return$_{t-7}$	0.004	-0.032	-0.060
Total Return$_{t-8}$	0.159*	0.141	0.133
PI_{t-1}	0.385*	0.136	0.095
PI_{t-2}	0.036	-0.199	-0.581***
PI_{t-3}	-0.023	0.020	-0.434**
PI_{t-4}	-0.063	0.153	-0.281
PI_{t-5}	-0.160	0.261	-0.165
PI_{t-6}	-0.523**	-0.170	-0.752***
PI_{t-7}	-0.425**	-0.803**	-0.926***
PI_{t-8}	0.206	0.081	-0.224
Spread	-19.369***	-16.774***	-15.995***
Term	-7.405***	-7.113***	-6.642***
Inflation	1.331	0.564	1.008
Constant	0.006	0.006	0.009
Adj. R^2	0.243	0.194	0.279
	Granger Causality		
Sentiment	0.030	0.224	0.005
Total Return	0.557	0.902	0.491

Note: This table shows the estimated coefficients from the VAR models with monthly total returns and the Pessimism Indicator. The set of the macroeconomic control variables includes the spread between Moody's Seasoned Baa- and Aaa-rated corporate bonds (*Spread*), the difference between the 10-year Treasury bond yield and the 3-month Treasury bill (*Term*), and the inflation rate of the U.S. (*Inflation*). All variables, except for total return, are considered as the first differences. Newey and West (1987) standard errors, which are robust to heteroscedasticity and autocorrelation, are used. Granger causality p-values are reported for both directions. * denotes significance at the 10% level, ** at the 5% level, and *** at the 1% level.

This indicates that investors might expect returns to decline when pessimism increases. Besides the short-term effects, all models find a long-term relationship between pessimism and total returns. For the dictionary approach and LSTM model, the sixth and seventh lag are significant at 5%, respectively, 1% level, and for the SVM model, the seventh lag is significant at the 5% level. These coefficients have the expected sign and the largest economic significance in the respective models. Therefore, the *PI* is an even better indicator than the *OI*.

The regression results for the *SQ*, as presented in Table 2.4, are consistent with the previous sentiment indicators but particularly noteworthy. The short-term and long-term relationships are very similar to the *PI*, with the long-term effects of sentiment again having the highest impact on total returns.

Table 2.4 – VAR Results: Monthly Lags Sentiment Quotient

	FTSE Nareit All REITs Total Return		
	SQ Dict.	**SQ SVM**	**SQ LSTM**
Total Return$_{t-1}$	-0.303**	-0.264**	-0.270**
Total Return$_{t-2}$	-0.063	-0.060	-0.081
Total Return$_{t-3}$	0.045	0.078	0.042
Total Return$_{t-4}$	0.054	0.053	-0.016
Total Return$_{t-5}$	0.134*	0.087	0.060
Total Return$_{t-6}$	-0.125*	-0.100	-0.074
Total Return$_{t-7}$	0.030	-0.033	-0.080
Total Return$_{t-8}$	0.143	0.151	0.111
SQ$_{t-1}$	-0.153**	-0.037	-0.154
SQ$_{t-2}$	-0.063	0.085	0.329**
SQ$_{t-3}$	0.028	0.071	0.258*
SQ$_{t-4}$	0.006	-0.038	0.161
SQ$_{t-5}$	0.029	-0.087	0.035
SQ$_{t-6}$	0.207**	0.153	0.450***
SQ$_{t-7}$	0.111*	0.426**	0.617***
SQ$_{t-8}$	-0.046	-0.008	0.123
Spread	-20.507***	-17.205***	-16.309***
Term	-6.810***	-7.353***	-6.874***
Inflation	1.308	0.552	0.940
Constant	0.006	0.006	0.008
Adj. R²	0.232	0.185	0.282
	Granger Causality		
Sentiment	0.051	0.297	0.004
Total Return	0.521	0.822	0.285

Note: This table shows the estimated coefficients from the VAR models with monthly total returns and the Sentiment Quotient. The set of the macroeconomic control variables includes the spread between Moody's Seasoned Baa- and Aaa-rated corporate bonds (*Spread*), the difference between the 10-year Treasury bond yield and the 3-month Treasury bill (*Term*), and the inflation rate of the U.S. (*Inflation*). All variables, except for total return, are considered as the first differences. Newey and West (1987) standard errors, which are robust to heteroscedasticity and autocorrelation, are used. Granger causality p-values are reported for both directions. * denotes significance at the 10% level, ** at the 5% level, and *** at the 1% level.

Since the results for the three sentiment indicators are consistent and reflected in the quarterly analysis presented in Appendix IV, it is reasonable to assume that there is a robust and well-founded relationship between them. The lower significance and magnitude of the short-term effects (one to three months ahead) lead us to interpret them as speculative reactions to observed changes in the stock market. Conversely, the more substantial long-term effects (six to seven

months ahead) could be seen as investors' reactions to changes in the underlying assets, i.e., the real estate market, which aligns with the idea of an initial market reaction that is followed by a delayed adjustment. The results regarding the long-term effects are consistent with previous research by Hausler et al. (2018) and Braun et al. (2020), who observed a similar lag structure between the media-expressed sentiment and direct real estate markets. The same applies when comparing our results to Ruscheinsky et al. (2018b) on news sentiment and REITs, who even find a similar structure in the coefficients of the sentiment variable. However, we find that social media sentiment is an even earlier indicator, by two months, than news sentiment.

In summary, the VAR results are highly consistent and robust, even when considering the quarterly regression. This supports our hypothesis that sentiment expressed on social media affects the returns of the REIT market. For all classifiers and indicators, we indeed observe a behavior that follows the theories of Grossman and Stiglitz (1976) and Hirshleifer and Teoh (2003). The results support the second hypothesis of social media sentiment being related to a short- and a long-term reaction in the REIT market. In this context, the immediate reaction of the market can be explained by the informed investors adjusting their behavior to the anticipated market development. This information is then transferred to uninformed investors through an information cascading and herd behavior effect, possibly leading to a delayed reaction. The delayed effect may be further strengthened through possible changes in the underlying real estate market, which cannot occur as immediately as financial market reactions, based on the asset's nature. All models demonstrate significant relationships, with the *PI* and *SQ* being the most informative indicators and accounting for a higher adjusted R^2.

When comparing the sentiment classifiers, the LSTM model is best at capturing the relationships and yields the highest adjusted R^2 for all sentiment indicators. However, contrary to our expectations, the dictionary approach outperforms the SVM model despite the previously described higher classification accuracy. We think that this is primarily due to two reasons. First, only a subset of the training dataset was used to train the model. Second, the dictionary approach might be better at classifying real estate-related messages because of the word lists applied. Hence, utilizing more training data or a real estate- related training dataset might have increased the SVM's performance. Ultimately, this partly supports our hypothesis that the more sophisticated a sentiment classifier is, the better the predictions.

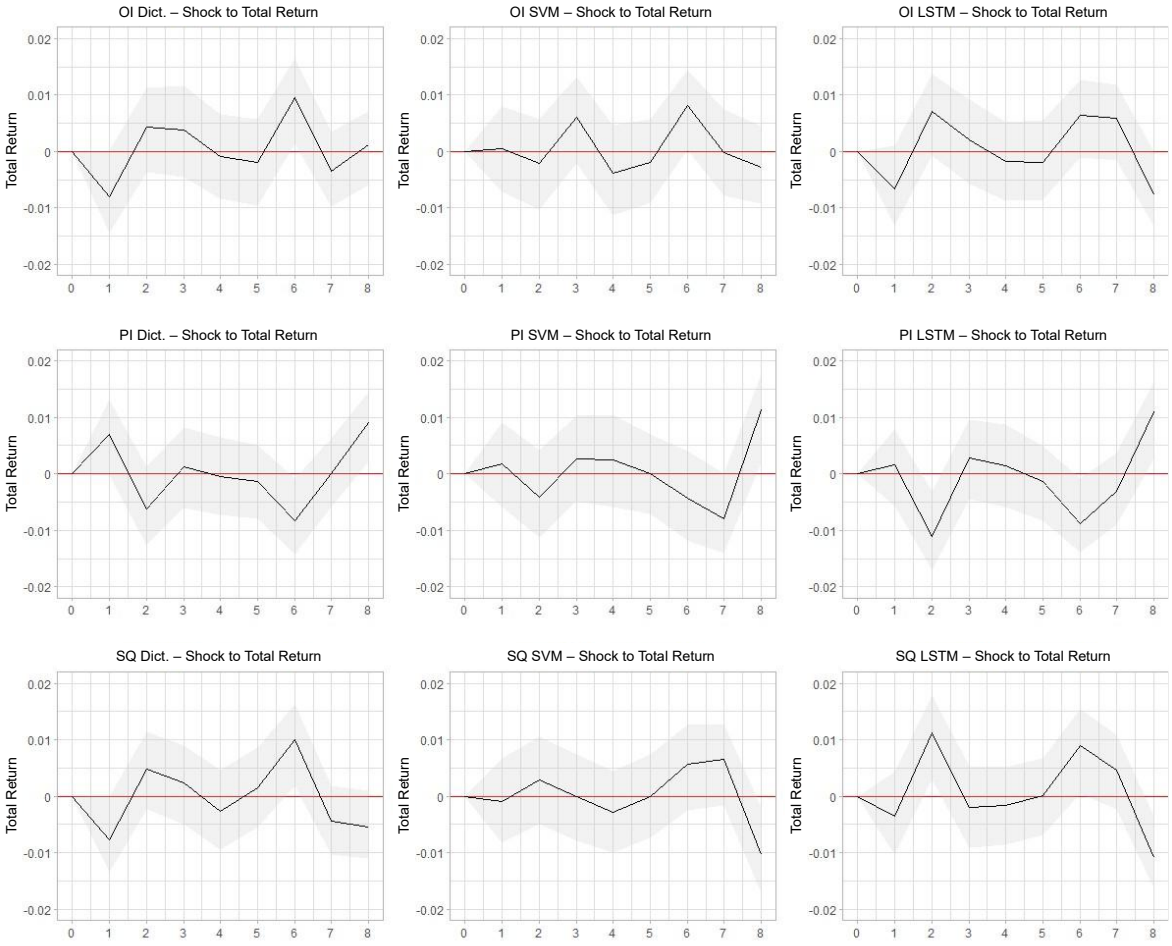
2.6.5 Granger Causality and Impulse Response Functions

Looking at the Granger causalities, these findings are confirmed. Tables 2.2 to 2.4 show that the sentiment indicators Granger-cause total returns, but not vice versa. Similar to the VAR results, the *OI* shows the lowest Granger causality, being significant at the 10% level only for the LSTM model. The dictionary approach and the LSTM model show significant Granger causality for the

PI and the *SQ*. For both indicators, the LSTM model is significant at the 1% level and is again the best predictor for future REIT market returns.

The results of the impulse response functions for all models are presented in Figure 2.8. Each line represents the outcomes for one of the sentiment indicators. Interestingly, the response of the total return appears to be independent of the classifier utilized. As anticipated, an increase in the *OI* and *SQ* leads to a positive response in total returns, while the opposite is true for the *PI*. These functions further support previous findings that sentiment effects can be categorized into short and long-term effects. The short-term effect occurs within the first three to four months and is typically characterized as a countermovement. Afterward, the long-term effect, which begins after six months, has a more significant and longer-lasting impact on the total return. Ultimately, the effect fades out in month eight.

Figure 2.8 – Impulse Response Functions



Note: The figure displays the impact of a one-time shock to the sentiment indicators on the total returns over time, as revealed by the impulse response functions. Each of the three sentiment indicators is represented by a separate line, while each column provides the results for one of the classification algorithms. The grey area represents the 95% confidence interval.

2.7 Conclusion

This study examines the impact of social media sentiment on REIT market returns from 2013 to 2022. Therefore, a total of 4.97 million tweets related to REITs were collected from Twitter and subjected to a rigorous selection process, ensuring the relevance of the remaining tweets and resulting in a final corpus of approximately 210,000 tweets. Those messages were classified as either positive, negative, or neutral and comprised in sentiment indices to compare them with the REIT market performance.

The results suggest a significant relationship between social media sentiment and total returns, regardless of whether the focus is on the percentage of optimistic or pessimistic messages or their ratio. As expected, an increase in optimism or in the Sentiment Quotient leads to higher returns in the subsequent months, while the opposite is true for pessimism. In line with prior research using newspapers or corporate disclosure, pessimism and the Sentiment Quotient have a greater impact than optimism on REIT returns. Nonetheless, the relationship for all sentiment indicators can be characterized by short- term effects, which are likely to be stock market driven, and long-term effects related to the underlying real estate market, with the latter having a greater impact. Hence, our first hypothesis that social media sentiment affects future REIT market returns is confirmed, which makes social media a valuable source of information for real estate investors. Furthermore, this behavior does support our second hypothesis of short and long- term reactions of the REIT market. Especially, with the growing public interest and the corresponding increase in message volume in recent years, future social media sentiment can be an even better indicator of public opinion.

This study employed three different sentiment classifiers to test the ability to understand the complex linguistic features inherent to social media messages. We compared a dictionary approach, designed to control for valence shifters and utilizing real estate word lists, with a support vector machine as a machine learning approach and a long short-term memory as a deep learning approach. Significant relationships of similar lag structure were found by all three models, which provides further evidence for our first hypothesis. Even the conventional and literature-approved dictionary approach seems to handle the linguistic complexity of social media messages and identify such relationships. The fact that the SVM and LSTM models find similar relationships suggests that it is not a coincidence.

Additionally, as it can be observed that the classification accuracy increases for these machine learning and deep learning approaches, our third hypothesis is confirmed. The LSTM approach finds the strongest relationship among all models, likely due to its better classification performance. The relationship detected by the SVM follows the same shape. However, it appears less significant, which might be related to its limited training. Increasing its training could improve

its classification accuracy, leading to more significant forecasts and elevating its performance above the dictionary approach used in this study. These findings partially confirm our fourth hypothesis that more sophisticated sentiment classifiers can better predict future REIT market returns. For real-world applications, the dictionary approach might be sufficient to get an initial understanding of the impact of social media; thereupon, investors may choose to use more sophisticated models that provide more accurate results for investment decisions.

These results are promising. Still, social media analysis is a relatively new field in real estate literature. Therefore, this paper offers a detailed selection process and comparison of sentiment classifiers that can be useful for future research and its application in practice. The selection process can be easily adjusted to identify relevant tweets for different topics using different search words or by replacing the GloVe with a topic modeling approach. Future research should focus on improving the selection process, for example, by applying different firm identifiers, signal words, or topic models. Even though it is more complex, advances in tweet selection may allow the approach to be applied to the direct real estate market. Regardless of the selection process, future research might consider increasing classification performance by using additional and more recent training data, finance-related training data, or by considering the impact of emojis. This will yield even better classification results and, combined with the increasing number of social media messages in recent years, provide additional possibilities for any analysis. Researchers could, for example, explore the quantity of messages as another sentiment indicator or change the dependent variable to measures such as stock price volatility or trading volume. In summary, our findings demonstrate the significant potential of social media data as a valuable source of information that merits further exploration.

2.8 Appendix

2.8.1 Appendix I – Support Vector Machine

Following Cortes and Vapnik (1995), a set of training vectors $(y_1, x_1), \dots, (y_l, x_l), y_l \in \{-1,1\}$ is separable by a hyperplane if the inequality $y_i(wx_i + b) \geq 1$ with $i = 1, \dots, l$ is fulfilled. The optimal hyperplane, representing the hyperplane that maximizes the distance between the separated points, is given by $w_0 z + b_0 = 0$. The maximized distance between the optimal hyperplane and the data points is given by:

$$p(w_o, b_o) = \frac{2}{|w_0|} = \frac{2}{\sqrt{w_0 * w_0}} \tag{2.7}$$

The optimal decision surface is determined by w_o, it can be written as a linear combination of training vectors:

$$w_o = \sum_{i=1}^{l} y_i \alpha_i^0 x_i \tag{2.8}$$

with $\alpha_i^0 \geq 0$. To find α_i, a quadratic programming function:

$$W(\Lambda) = \Lambda^T 1 - \frac{1}{2}\Lambda^T D \Lambda \tag{2.9}$$

with $\Lambda^T = (\alpha_1, \dots, \alpha_l)$, and the constraints $\Lambda \geq 0; \Lambda^T Y = 0$, where $1^T = (1, \dots, 1)$ is an l-dimensional vector, $Y^T = (y_1, \dots, y_l)$ is an l-dimensional vector of labels, and D is a symmetric lxl matrix $D_i j = y_i y_j x_i \cdot x_j$ with $i, j = 1, \dots, l$, needs to be solved. As the formula for w_0 is given, it can be solved for b_0, which provides all parameters for stating the optimal hyperplane.

Cortes and Vapnik (1995) state that the optimal hyperplane may still not be found if the data is not separable without any errors. Therefore, the soft margin classifier is introduced to solve this limitation. A new parameter C is introduced, modifying the quadratic programming function, given in equation (2.9), to allow for some errors:

$$W(\Lambda) = \Lambda^T 1 - \frac{1}{2}\left[\Lambda^T D \Lambda + \frac{\delta^2}{C}\right] \tag{2.10}$$

where $\delta = \alpha_{\max} = \max(\alpha_1, \dots, \alpha_l)$. This addition enables the SVM to find the optimal hyperplane for the remaining data.

New data, thus unknown vectors, are classified by applying the sign function. Therefore, the output is binary with a value of 1 or -1 for the classification categories. In the context of this study, the SVM outputs a value of 1 for positive and -1 for negative messages when classifying the input data. As the output of an SVM hence is not probabilistic, Platt scaling is applied to get

probabilistic output values that enable the desired sentiment classification. Platt (1999) uses a parametric form of a sigmoid:

$$P(y = 1 \mid f) = \frac{1}{1 + \exp(Af + B)} \tag{2.11}$$

The variables A & B are fit by a maximum likelihood estimation based on a training set (f_i, y_i). As stated by Platt (1999), to find $A\&B$, a new training set (f_i, t_i) is defined. With t_i as the target probabilities:

$$t_i = \frac{y_i + 1}{2} \tag{2.12}$$

Finally, to calculate the parameters $A\&B$, the negative log-likelihood function of the training data must be minimized:

$$\min - \sum_i t_i \log(p_i) + (1 - t_i)\log(1 - p_i) \tag{2.13}$$

with

$$p_i = \frac{1}{1 + \exp(Af_i + B)} \tag{2.14}$$

After the scaling is applied, the probability of a tweet having a positive sentiment is presented, enabling the desired classification.

2.8.2 Appendix II – Long Short-Term Memory

Facing the shortcomings of conventional RNNs, the LSTM network adds a memory cell c_j, which evolves during the constant error carousel (CEC). The CEC is a mechanism where the cell state, the value saved in the memory cell, is multiplied by a weight of 1. The cell state can only be changed by values allowed to pass through the input gate, and the output gate regulates access to the memory cell and protects other units from interference (Hochreiter & Schmidhuber, 1997). The output of the memory cell can be stated as follows:

$$y^{c_j}(t) = y^{out_j}(t) h\left(s_{c_j}(t)\right) \tag{2.15}$$

y^{out_j} describes the activation of the output gate. The output gate and the input gate y^{in_j} are given as:

$$y^{in_j}(t) = f_{\text{in}_j}\left(\text{net}_{in_j}(t)\right); y^{out_j}(t) = f_{out_j}\left(\text{net}_{out_j}(t)\right) \tag{2.16}$$

where

$$\text{net}_{in_j}(t) = \sum_u w_{in_j u} y^u (t-1);\ \text{net}_{out_j}(t) = \sum_u w_{out_j u} y^u (t-1) \tag{2.17}$$

with net_j being the current net input for unit j. Modern LSTM implementations add an additional gate to the memory cell called the forget gate. The forget gate was introduced by Gers et al. (2000) as an addition to the memory cell. The authors mention that, despite the presence of the CEC, the internal state can still grow unchecked and pose a threat to the stability of the network, which is countered by an additional gate unit. The forget gate acts as a reset for the LSTM cell, it is forgetting its internal information. As stated in Gers et al. (2000), the activation of the forget gate y^{φ} states as:

$$y^{\varphi j}(t) = f_{\varphi j}\left(\text{net}_{\varphi j}(t)\right) \tag{2.18}$$

with $net_{\varphi j}$ being the net input:

$$\text{net}_{\varphi j}(t) = \sum_u w_{\varphi j u} y^u (t-1) \tag{2.19}$$

Gers et al. (2000) further adjusted the calculation formula for the internal cell state s_{c_j} by adding the forget gate:

$$s_{c_j}(0) = 0; s_{c_j}(t) = y^{\varphi j}(t) s_{c_j}(t-1) + y^{in_j}(t) g\left(\text{net}_{c_j}(t)\right) \text{ for } t > 0 \tag{2.20}$$

With the addition of the forget gate, the LSTM unit now includes three gates, determining when different values can flow through it. The output given by the three gates ranges between 0 and 1. Soutner and Müller (2013) describe how the different gates can influence the LSTM cell: The input gates can decide how an input given can influence the cell. With a gate value close to 0, the input does not change the cells' internal values. The forget gate can further decide if stored information in the cell is kept or reset by giving a value close to 1 for keeping it and 0 to reset it. The output gate further decides when the information stored in the unit should be put out.

2.8.3 Appendix III – Signal Words

Table 2.5 – Signal Words

Start word	Signal words
REIT	reit; reits; estate; portfolio; retail; sector; stock; real; property; industrial; dividend; company; yield; focus; business; exposure; favorite; consider; growth; pick; office; healthcare; medical; market; investor

2.8.4 Appendix IV – Vector Autoregression: Quarterly Lags

Table 2.6 – VAR Results: Quarterly Lags Optimism Indicator

	FTSE Nareit All REITs Total Return		
	OI Dict.	**OI SVM**	**OI LSTM**
Total Return$_{t-1}$	-0.618***	-0.668***	-0.652***
Total Return$_{t-2}$	-0.389*	-0.471***	-0.469***
Total Return$_{t-3}$	-0.130	-0.135	-0.196
Total Return$_{t-4}$	-0.109	-0.162**	-0.126
OI$_{t-1}$	-0.257	0.218	0.157
OI$_{t-2}$	0.078	0.470**	0.717
OI$_{t-3}$	-0.896	-0.197	-0.123
OI$_{t-4}$	-0.818	0.739**	0.192
Spread	-26.972***	-28.952***	-30.201***
Term	-5.430	-3.709	-3.279
Inflation	8.766	11.552	9.411**
Constant	-0.001	-0.001	-0.002
Adj. R²	0.497	0.555	0.532
	Granger Causality		
Sentiment	0.819	0.319	0.490
Total Return	0.500	0.545	0.002

Note: This table shows the estimated coefficients from the VAR models with quarterly total returns and the Optimism Indicator. The set of the macroeconomic control variables includes the spread between Moody's Seasoned Baa- and Aaa-rated corporate bonds (*Spread*), the difference between the 10-year Treasury bond yield and the 3-month Treasury bill (*Term*), and the inflation rate of the U.S. (*Inflation*). All variables are considered as second differences, except for total returns, which are first differences. Newey and West (1987) standard errors, which are robust to heteroscedasticity and autocorrelation, are used. Granger causality p-values are reported for both directions. * denotes significance at the 10% level, ** at the 5% level, and *** at the 1% level.

Table 2.7 – VAR Results: Quarterly Lags Pessimism Indicator

	FTSE Nareit All REITs Total Return		
	PI Dict.	**PI SVM**	**PI LSTM**
Total Return$_{t-1}$	-0.646***	-0.575***	-0.660***
Total Return$_{t-2}$	-0.404**	-0.364*	-0.563***
Total Return$_{t-3}$	-0.142	-0.215*	-0.220
Total Return$_{t-4}$	-0.085	-0.248	-0.167
PI$_{t-1}$	-0.258	1.633	-0.296
PI$_{t-2}$	-1.310**	1.280	-1.457***
PI$_{t-3}$	-0.535	0.630	-0.222
PI$_{t-4}$	-0.422	0.096	-0.753
Spread	-31.831**	-22.583*	-27.135***
Term	-3.801	-5.658	-2.602
Inflation	8.419	8.521	11.188**
Constant	-0.001	-0.002	-0.002
Adj. R²	0.513	0.488	0.542
	Granger Causality		
Sentiment	0.669	0.897	0.409
Total Return	0.362	0.253	0.062

Note: This table shows the estimated coefficients from the VAR models with quarterly total returns and the Pessimism Indicator. The set of the macroeconomic control variables includes the spread between Moody's Seasoned Baa- and Aaa-rated corporate bonds (*Spread*), the difference between the 10-year Treasury bond yield and the 3-month Treasury bill (*Term*), and the inflation rate of the U.S. (*Inflation*). All variables are considered as second differences, except for total returns, which are first differences. Newey and West (1987) standard errors, which are robust to heteroscedasticity and autocorrelation, are used. Granger causality p-values are reported for both directions. * denotes significance at the 10% level, ** at the 5% level, and *** at the 1% level.

Table 2.8 – VAR Results: Quarterly Lags Sentiment Quotient

	FTSE Nareit All REITs Total Return		
	SQ Dict.	**SQ SVM**	**SQ LSTM**
Total Return$_{t-1}$	-0.606***	-0.621***	-0.649***
Total Return$_{t-2}$	-0.354*	-0.424**	-0.549***
Total Return$_{t-3}$	-0.091	-0.217*	-0.222
Total Return$_{t-4}$	-0.028	-0.233	-0.165
SQ$_{t-1}$	0.116	-0.586	0.143
SQ$_{t-2}$	0.582**	-0.270	0.932**
SQ$_{t-3}$	0.202	-0.262	0.043
SQ$_{t-4}$	0.107	0.061	0.489
Spread	-36.118***	-23.082	-27.454***
Term	-3.452	-4.747	-2.694
Inflation	7.647	9.992*	11.169**
Constant	-0.001	-0.002	-0.002
Adj. R^2	0.543	0.484	0.546
	Granger Causality		
Sentiment	0.402	0.921	0.381
Total Return	0.438	0.126	0.028

Note: This table shows the estimated coefficients from the VAR models with quarterly total returns and the Sentiment Quotient. The set of the macroeconomic control variables includes the spread between Moody's Seasoned Baa- and Aaa-rated corporate bonds (*Spread*), the difference between the 10-year Treasury bond yield and the 3-month Treasury bill (*Term*), and the inflation rate of the U.S. (*Inflation*). All variables are considered as second differences, except for total returns, which are first differences. Newey and West (1987) standard errors, which are robust to heteroscedasticity and autocorrelation, are used. Granger causality p-values are reported for both directions. * denotes significance at the 10% level, ** at the 5% level, and *** at the 1% level.

2.9 References

Ankit & Saleena, N. (2018). An ensemble classification system for Twitter sentiment analysis. *Procedia Computer Science*, *132*, 937–946. https://doi.org/10.1016/j.procs.2018.05.109

Antweiler, W., & Frank, M. Z. (2004). Is all that talk just noise? The information content of internet stock message boards. *The Journal of Finance*, *59*(3), 1259–1294. https://doi.org/10.1111/j.1540-6261.2004.00662.x

Ba, J. L., Kiros, J. R., & Hinton, G. E. (2016). *Layer normalization*. arXiv. https://doi.org/10.48550/arXiv.1607.06450

Bahrainian, S.-A., & Dengel, A. (2013). Sentiment analysis using sentiment features. In *2013 IEEE/WIC/ACM International Joint Conferences on Web Intelligence (WI) and Intelligent Agent Technologies (IAT)* (pp. 26–29). IEEE. https://doi.org/10.1109/WI-IAT.2013.145

Baker, M., & Wurgler, J. (2007). Investor sentiment in the stock market. *Journal of Economic Perspectives*, *21*(2), 129–151. https://doi.org/10.1257/jep.21.2.129

Beracha, E., & Downs, D. H. (2015). Value and momentum in commercial real estate: A market-level analysis. *The Journal of Portfolio Management*, *41*(6), 48–61. https://doi.org/10.3905/jpm.2015.41.6.048

Beracha, E., Lang, M., & Hausler, J. (2019). On the relationship between market sentiment and commercial real estate performance—A textual analysis examination. *Journal of Real Estate Research*, *41*(4), 605–638. https://doi.org/10.22300/0896-5803.41.4.605

Bollen, J., Mao, H., & Zeng, X. (2011). Twitter mood predicts the stock market. *Journal of Computational Science*, *2*(1), 1–8. https://doi.org/10.1016/j.jocs.2010.12.007

Braun, J., Hausler, J., & Schäfers, W. (2020). Artificial intelligence, news sentiment, and property market liquidity. *Journal of Property Investment & Finance*, *38*(4), 309–325. https://doi.org/10.1108/JPIF-08-2019-0100

Broersma, M., & Graham, T. (2013). Twitter as a news source. *Journalism Practice*, *7*(4), 446–464. https://doi.org/10.1080/17512786.2013.802481

Brooks, C., & Tsolacos, S. (2010). *Real estate modelling and forecasting*. Cambridge University Press.

Carstens, R., & Freybote, J. (2019). Tone in REIT financial statements and institutional investments. *Journal of Property Research*, *36*(3), 227–244. https://doi.org/10.1080/09599916.2019.1650802

Carstens, R., & Freybote, J. (2021). Can the textual tone in REIT financial statements improve the information environment for commercial real estate investors? An investigation. *Journal of Real Estate Research*, *43*(3), 335–354. https://doi.org/10.1080/08965803.2021.1985923

Chen, H., De, P., Hu, Y., & Hwang, B.-H. (2011). Sentiment revealed in social media and its effect on the stock market. *In 2011 IEEE Statistical Signal Processing Workshop (SSP)* (pp. 25–28). IEEE. https://doi.org/10.1109/SSP.2011.5967675

Clayton, J., Ling, D. C., & Naranjo, A. (2009). Commercial real estate valuation: Fundamentals versus investor sentiment. *The Journal of Real Estate Finance and Economics*, *38*(1), 5–37. https://doi.org/10.1007/s11146-008-9130-6

Cortes, C., & Vapnik, V. (1995). Support-vector networks. *Machine Learning*, *20*(3), 273–297. https://doi.org/10.1007/BF00994018

Da, Z., Engelberg, J., & Gao, P. (2011). In search of attention. *The Journal of Finance*, *66*(5), 1461–1499. https://doi.org/10.1111/j.1540-6261.2011.01679.x

Das, P., Freybote, J., & Marcato, G. (2015a). An investigation into sentiment-induced institutional trading behavior and asset pricing in the REIT market. *The Journal of Real Estate Finance and Economics*, *51*(2), 160–189. https://doi.org/10.1007/s11146-014-9490-z

Das, P., Ziobrowski, A., & Coulson, N. E. (2015b). Online information search, market fundamentals and apartment real estate. *The Journal of Real Estate Finance and Economics*, *51*(4), 480–502. https://doi.org/10.1007/s11146-015-9496-1

Doran, J. S., Peterson, D. R., & Price, S. M. (2012). Earnings conference call content and stock price: The case of REITs. *The Journal of Real Estate Finance and Economics*, *45*(2), 402–434. https://doi.org/10.1007/s11146-010-9266-z

Dorle, S., & Pise, N. (2018). Political sentiment analysis through social media. In *2018 Second International Conference on Computing Methodologies and Communication (ICCMC)* (pp. 869–873). IEEE. https://doi.org/10.1109/ICCMC.2018.8487879

Freybote, J. (2016). Real estate sentiment as information for REIT bond pricing. *Journal of Property Research*, *33*(1), 18–36. https://doi.org/10.1080/09599916.2016.1146791

Freybote, J., & Seagraves, P. A. (2017). Heterogeneous investor sentiment and institutional real estate investments. *Real Estate Economics*, *45*(1), 154–176. https://doi.org/10.1111/1540-6229.12132

Fu, X., Yang, J., Li, J., Fang, M., & Wang, H. (2018). Lexicon-enhanced LSTM with attention for general sentiment analysis. *IEEE Access*, *6*, 71884–71891. https://doi.org/10.1109/ACCESS.2018.2878425

Gallimore, P., & Gray, A. (2002). The role of investor sentiment in property investment decisions. *Journal of Property Research*, *19*(2), 111–120. https://doi.org/10.1080/09599910110110671

Gers, F. A., Schmidhuber, J., & Cummins, F. (2000). Learning to forget: Continual prediction with LSTM. *Neural Computation*, *12*(10), 2451–2471. https://doi.org/10.1162/089976600300015015

Go, A., Bhayani, R., & Huang, L. (2009). *Twitter sentiment classification using distant supervision* [CS224N project report]. Stanford University. https://www-cs.stanford.edu/people/alecmgo/papers/TwitterDistantSupervision09.pdf

Granger, C. W. J. (1969). Investigating causal relations by econometric models and cross-spectral methods. *Econometrica*, *37*(3), 424–438. https://doi.org/10.2307/1912791

Grossman, S. J., & Stiglitz, J. E. (1976). Information and competitive price systems. *American Economic Review*, *66*(2), 246–253. http://www.jstor.org/stable/1817229

Gupta, R., & Chen, M. (2020). Sentiment analysis for stock price prediction. In *2020 IEEE Conference on Multimedia Information Processing and Retrieval (MIPR)* (pp. 213–218). IEEE. https://doi.org/10.1109/MIPR49039.2020.00051

Hannum, C., Arslanli, K. Y., & Kalay, A. F. (2019). Spatial analysis of Twitter sentiment and district-level housing prices. *Journal of European Real Estate Research*, *12*(2), 173–189. https://doi.org/10.1108/JERER-08-2018-0036

Hausler, J., Ruscheinsky, J., & Lang, M. (2018). News-based sentiment analysis in real estate: A machine learning approach. *Journal of Property Research*, *35*(4), 344–371. https://doi.org/10.1080/09599916.2018.1551923

Heinig, S., Nanda, A., & Tsolacos, S. (2020). Which sentiment indicators matter? Evidence from the European Commercial Real Estate Market. *Journal of Real Estate Research*, *42*(4), 499–530. https://doi.org/10.1080/08965803.2020.1845562

Henry, E., & Leone, A. J. (2016). Measuring qualitative information in capital markets research: Comparison of alternative methodologies to measure disclosure tone. *The Accounting Review*, *91*(1), 153–178. https://doi.org/10.2308/accr-51161

Hirshleifer, D., & Teoh, S. H. (2003). Herd behaviour and cascading in capital markets: A review and synthesis. *European Financial Management*, *9*(1), 25–66. https://doi.org/10.1111/1468-036X.00207

Hochreiter, S. (1991). *Untersuchungen zu dynamischen neuronalen Netzen* [Diploma thesis]. Technische Universität München. https://www.bioinf.jku.at/publications/older/3804.pdf

Hochreiter, S., & Schmidhuber, J. (1997). Long short-term memory. *Neural Computation*, *9*(8), 1735–1780. https://doi.org/10.1162/neco.1997.9.8.1735

Hoesli, M., Lizieri, C., & MacGregor, B. (2008). The inflation hedging characteristics of US and UK investments: A multi-factor error correction approach. *The Journal of Real Estate Finance and Economics*, *36*(2), 183–206. https://doi.org/10.1007/s11146-007-9062-6

Hohenstatt, R., Käsbauer, M., & Schäfers, W. (2011). "Geco" and its potential for real estate research: Evidence from the U.S. housing market. *Journal of Real Estate Research*, *33*(4), 471–506. https://doi.org/10.1080/10835547.2011.12091318

Jin, Z., Yang, Y., & Liu, Y. (2020). Stock closing price prediction based on sentiment analysis and LSTM. *Neural Computing and Applications*, *32*(13), 9713–9729. https://doi.org/10.1007/s00521-019-04504-2

Jones, R. (2022). *Internet & text slang dictionary*. Retrieved August 22, 2022, from https://www.noslang.com/dictionary/

Keynes, J. M. (1936). *The general theory of employment, interest, and money*. Macmillan.

Li, F. (2010). The information content of forward-looking statements in corporate filings – A Naïve Bayesian machine learning approach. *Journal of Accounting Research*, *48*(5), 1049–1102. https://doi.org/10.1111/j.1475-679X.2010.00382.x

Li, T., van Dalen, J., & van Rees, P. J. (2018). More than just noise? Examining the information content of stock microblogs on financial markets. *Journal of Information Technology*, *33*(1), 50–69. https://doi.org/10.1057/s41265-016-0034-2

Lin, C. Y., Rahman, H., & Yung, K. (2009). Investor sentiment and REIT returns. *The Journal of Real Estate Finance and Economics*, *39*(4), 450–471. https://doi.org/10.1007/s11146-008-9115-5

Ling, D. C., Naranjo, A., & Scheick, B. (2014). Investor sentiment, limits to arbitrage and private market returns. *Real Estate Economics*, *42*(3), 531–577. https://doi.org/10.1111/1540-6229.12037

Loughran, T., & McDonald, B. (2011). When is a liability not a liability? Textual analysis, dictionaries, and 10-Ks. *The Journal of Finance*, *66*(1), 35–65. https://doi.org/10.1111/j.1540-6261.2010.01625.x

Manning, C. D., Raghavan, P., & Schütze, H. (2008). The term vocabulary and postings lists. In *Introduction to information retrieval* (pp. 18–44). Cambridge University Press. https://doi.org/10.1017/CBO9780511809071

Marcato, G., & Nanda, A. (2016). Information content and forecasting ability of sentiment indicators: Case of real estate market. *Journal of Real Estate Research*, *38*(2), 165–204. https://doi.org/10.1080/10835547.2016.12091442

Maynard, D., Bontcheva, K., & Rout, D. (2012). Challenges in developing opinion mining tools for social media. In *Proceedings of @NLP can u tag #user_generated_content?! Workshop at LREC 2012* (pp. 15–22). European Language Resources Association. http://www.lrec-conf.org/proceedings/lrec2012/workshops/21.LREC2012%20NLP4UGC%20Proceedings.pdf

Nareit. (2022). FTSE Nareit All REITs Index [Fact sheet]. https://www.ftserussell.com/products/indices/nareit

Newey, W. K., & West, K. D. (1987). A simple, positive semi-definite, heteroskedasticity and autocorrelation consistent covariance matrix. *Econometrica*, *55*(3), 703–708. https://doi.org/10.2307/1913610

Nowak, A., & Smith, P. (2017). Textual analysis in real estate. *Journal of Applied Econometrics*, *32*(4), 896–918. https://doi.org/10.1002/jae.2550

Paulus, N. M., Koelbl, M., & Schaefers, W. (2022). Can textual analysis solve the underpricing puzzle? A US REIT study. *Journal of Property Investment & Finance*, *40*(6), 548–570. https://doi.org/10.1108/JPIF-06-2021-0052

Paulussen, S., & Harder, R. A. (2014). Social media references in newspapers. *Journalism Practice*, *8*(5), 542–551. https://doi.org/10.1080/17512786.2014.894327

Pennington, J., Socher, R., & Manning, C. (2014). GloVe: Global vectors for word representation. In A. Moschitti, B. Pang, & W. Daelemans (Eds.), *Proceedings of the 2014 Conference on Empirical Methods in Natural Language Processing (EMNLP)* (pp. 1532–1543). Association for Computational Linguistics. https://doi.org/10.3115/v1/D14-1162

Platt, J. C. (1999). Probabilistic outputs for support vector machines and comparisons to regularized likelihood methods. In A. J. Smola, P. Bartlett, B. Schölkopf, & D. Schuurmans (Eds.), *Advances in large margin classifiers* (pp. 61–74). MIT Press.

Plößl, F., Paulus, N. M., & Just, T. (2023). Trade vs. daily press: The role of news coverage and sentiment in real estate market analysis. *Journal of Property Research*, *40*(4), 333–364. https://doi.org/10.1080/09599916.2023.2199764

Price, S. M., Seiler, M. J., & Shen, J. (2017). Do investors infer vocal cues from CEOs during quarterly REIT conference calls? *The Journal of Real Estate Finance and Economics*, *54*(4), 515–557. https://doi.org/10.1007/s11146-016-9557-0

Rinker, T. W. (2019). *Sentiment: Calculate text polarity sentiment* (Version 2.7.1) [R package]. https://github.com/trinker/sentimentr

Rozin, P., & Royzman, E. B. (2001). Negativity bias, negativity dominance, and contagion. *Personality and Social Psychology Review*, *5*(4), 296–320. https://doi.org/10.1207/S15327957PSPR0504_2

Ruscheinsky, J. R., Kandlbinder, K., Dietzel, M. A., Rochdi, K., & Schaefers, W. (2018a). Forecasting the German real estate market with dictionary-based textual analysis. In *25th Annual European Real Estate Society Conference*. European Real Estate Society. https://doi.org/10.15396/eres2018_198

Ruscheinsky, J. R., Lang, M., & Schäfers, W. (2018b). Real estate media sentiment through textual analysis. *Journal of Property Investment & Finance*, *36*(5), 410–428. https://doi.org/10.1108/JPIF-07-2017-0050

Sarker, I. H. (2021). Deep learning: A comprehensive overview on techniques, taxonomy, applications and research directions. *SN Computer Science*, *2*(6), 420. https://doi.org/10.1007/s42979-021-00815-1

Shi, Y., Zheng, Y., Guo, K., & Ren, X. (2021). Stock movement prediction with sentiment analysis based on deep learning networks. *Concurrency and Computation: Practice and Experience*, *33*(6), Article e6076. https://doi.org/10.1002/cpe.6076

Shiller, R. J. (2017). Narrative economics. *American Economic Review*, *107*(4), 967–1004. https://doi.org/10.1257/aer.107.4.967

Shiller, R. J. (2020). *Narrative economics: How stories go viral and drive major economic events*. Princeton University Press.

Soo, C. K. (2015). *Quantifying animal spirits: News media and sentiment in the housing market* (Working Paper No. 1200). Ross School of Business. https://doi.org/10.2139/ssrn.2330392

Soutner, D., & Müller, L. (2013). Application of LSTM neural networks in language modelling. In I. Habernal & V. Matoušek (Eds.), *Text, speech, and dialogue* (Vol. 8082, pp. 105–112). Springer. https://doi.org/10.1007/978-3-642-40585-3_14

Sprenger, T. O., Tumasjan, A., Sandner, P. G., & Welpe, I. M. (2014). Tweets and trades: The information content of stock microblogs. *European Financial Management*, *20*(5), 926–957. https://doi.org/10.1111/j.1468-036X.2013.12007.x

Tan, M. J., & Guan, C. (2021). Are people happier in locations of high property value? Spatial temporal analytics of activity frequency, public sentiment and housing price using Twitter

data. *Applied Geography*, *132*, Article 102474. https://doi.org/10.1016/j.apgeog.2021.102474

Tetlock, P. C. (2007). Giving content to investor sentiment: The role of media in the stock market. *The Journal of Finance*, *62*(3), 1139–1168. https://doi.org/10.1111/j.1540-6261.2007.01232.x

Walker, C. B. (2014). Housing booms and media coverage. *Applied Economics*, *46*(32), 3954–3967. https://doi.org/10.1080/00036846.2014.948675

Wang, C., Wang, T., & Yuan, C. (2020). Does applying deep learning in financial sentiment analysis lead to better classification performance? *Economics Bulletin*, *40*(2), 1091–1105.

Watanabe, K., & Zhou, Y. (2020). Theory-driven analysis of large corpora: Semisupervised topic classification of the UN speeches. *Social Science Computer Review*, *40*(2), 346–366. https://doi.org/10.1177/0894439320907027

Wu, S., Liu, Y., Zou, Z., & Weng, T.-H. (2022). S_I_LSTM: Stock price prediction based on multiple data sources and sentiment analysis. *Connection Science*, *34*(1), 44–62. https://doi.org/10.1080/09540091.2021.1940101

Xu, Y., & Cohen, S. B. (2018). Stock movement prediction from tweets and historical prices. In I. Gurevych & Y. Miyao (Eds.), *Proceedings of the 56th Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers)* (pp. 1970–1979). Association for Computational Linguistics. https://doi.org/10.18653/v1/P18-1183

Yadav, A., & Vishwakarma, D. K. (2020). Sentiment analysis using deep learning architectures: A review. *Artificial Intelligence Review*, *53*(6), 4335–4385. https://doi.org/10.1007/s10462-019-09794-5

Yildirim, S., Jothimani, D., Kavaklioglu, C., & Basar, A. (2019). Deep learning approaches for sentiment analysis on financial microblog dataset. In *2019 IEEE International Conference on Big Data (Big Data)* (pp. 5581–5584). IEEE. https://doi.org/10.1109/BigData47090.2019.9006056

Zamani, M., & Schwartz, H. A. (2017). Using Twitter language to predict the real estate market. In M. Lapata, P. Blunsom, & A. Koller (Eds.), *Proceedings of the 15th Conference of the European Chapter of the Association for Computational Linguistics: Volume 2, Short Papers* (pp. 28–33). Association for Computational Linguistics. https://doi.org/10.18653/v1/E17-2005

Zhou, X., Tao, X., Rahman, M. M., & Zhang, J. (2017). Coupling topic modelling in opinion mining for social media analysis. In *WI '17: Proceedings of the International Conference on*

Web Intelligence (pp. 533–540). Association for Computing Machinery. https://doi.org/10.1145/3106426.3106459

3 Fundamentals or Noise? The Informative Value of Twitter Sentiment for REIT Returns

3.1 Abstract

Twitter sentiment has been found to predict REIT returns, which raises the question as to whether this is due to it containing fundamental information or merely representing noise. We hypothesize that the informative value of Twitter sentiment for REIT returns changed due to an increased engagement of individual investors with REITs during the COVID-19 pandemic. We extract sentiment measures from REIT-related tweets over the period of 2014 to 2022 and assess 1) whether the explanatory power of fundamentals for Twitter sentiment differs between the pre- and COVID period and 2) whether the relationship of sentiment and REIT returns is driven by rational (fundamentals-based) or irrational sentiment. We find that fundamentals such as inflation, stock market returns, credit risk spread, individual stock investor sentiment, and volatility have more predictive power for optimistic, pessimistic, and overall Twitter sentiment in the COVID period than in the pre-COVID period. Furthermore, pessimistic Twitter sentiment has the highest explanatory power for REIT returns, with irrational (rational) pessimistic Twitter sentiment predicting REIT returns in the pre-COVID (COVID) period. Our findings provide evidence for a structural break in the informative value of Twitter sentiment for REIT returns.

Keywords: Real Estate Investment Trusts, Investor Sentiment, Individual Investors, Social Media, Textual Analysis

3.2 Introduction

Over the last decade, social media platforms such as Twitter, Seeking Alpha, and StockTwits have increased in importance for the stock market investment community, particularly for individual (unsophisticated) investors (Eliner & Kobilov, 2023). Social media allows investors to observe, talk to as well as share their ideas and strategies with each other, which then can impact investment decisions (Han et al., 2022). This has led to a new paradigm, which Hirshleifer (2020) refers to as "Social Economics and Finance". It is distinct from paradigms such as behavioral finance or information economics and focuses on the impact of social media interactions on decision-making and investor behavior in financial markets. Understanding social economics and finance is important as social interactions of investors can lead to, amongst others, swings in economic sentiment and bubbles (social transmission bias; Hirshleifer, 2020).

Paulus et al. (2025) find that Twitter sentiment allows for predicting aggregated REIT returns, which is in line with non-REIT studies that find social media sentiment to have informative value for stock returns, trading volume, and stock market liquidity (Ammari et al., 2023; Gu & Kurov, 2020; Karampatsas et al., 2023; Ranco et al., 2015; Sul et al., 2017; Tan & Tas, 2021). However, Paulus et al. (2025) fail to explain why Twitter sentiment predicts REIT returns. On the one hand, Twitter sentiment could reflect fundamentals that affect commercial real estate and REIT markets (rational or fundamentals-based sentiment). On the other hand, it could reflect irrationality or noise that impacts investor decision-making and REIT returns (irrational sentiment). This study uses Paulus et al. (2025) as a starting point to investigate what information Twitter sentiment contains and why it can explain REIT returns.

The REIT industry is uniquely suited for our investigation as it allows for a quasi-experimental design for the following reasons: Twitter users tend to focus on large-cap stocks (Cookson et al., 2024) and popular stocks they are familiar with, such as Facebook, Netflix, or Tesla (Welch, 2022). Furthermore, Twitter is a social media platform used rather for leisure comments than professional ones and more likely reflects users as consumers (Tang, 2018). Considering the REIT industry is less consumer-focused and highly specialized, its ability to attract the attention of a large share of unsophisticated investors on Twitter is likely limited. In fact, Zeitun et al. (2023) find that real estate industry-level returns have less sensitivity to changes in Twitter sentiment than returns of industries such as healthcare, consumer discretionary, and information technology.

However, individual investors gravitate toward attention-grabbing stocks (Barber & Odean, 2008), and during the pandemic, stocks that received a lot of COVID-related media attention were traded more by these investors (Ozik et al., 2021). Considering that commercial real estate markets were heavily affected by COVID, ranging from a negative impact (e.g., office, hospitality)

to a positive one (e.g., logistics), and the situation of landlords was extensively discussed in the media, more individual investors likely focused on REITs during the pandemic than before. Furthermore, the COVID period in the U.S. was characterized by stay-at-home orders, relief cash payments to households, and FinTech innovations such as Robinhood with commission-free trading, which increased the time individual investors were able to engage on social media and invest excess funds in the stock market. Welch (2022) finds that small retail investors using Robinhood collectively stayed calm at the beginning of the COVID period and purchased more stocks. Ozik et al. (2021) show that individual (retail) investors traded more on Robinhood when lockdowns went into effect. Furthermore, REITs were not only likely to get more attention from individual investors during the pandemic, but these investors also had more time to obtain information about the REIT industry and individual firms.

Consequently, we expect a structural break in the explanatory power of economic, capital, and real estate market fundamentals for REIT-specific Twitter sentiment as well as the informative value of Twitter sentiment for REIT returns between the pre-COVID and COVID period. In particular, as our first hypothesis, we expect that fundamentals are able to predict Twitter sentiment to a higher degree in the COVID period than in the pre-COVID period. This follows the assumption that individual investors had more time and possibilities to access valuable information due to worldwide lockdowns. Furthermore, we hypothesize that the relationship of Twitter sentiment and REIT returns in the pre-COVID period is driven by noise or irrationality (irrational Twitter sentiment) while it was driven by the rational (fundamentals-based) component of Twitter sentiment in the COVID period.

Using natural language processing and Twitter data over the period of 2014 to 2022, we follow Paulus et al. (2025) to extract optimistic, pessimistic, and overall sentiment from tweets about REITs. Next, we investigate the relationship of each sentiment measure with macroeconomic, capital market, and commercial real estate market variables to assess which type of fundamental information affects REIT-related Twitter sentiment. In the last step, we separate Twitter sentiment into a rational and irrational sentiment component to analyze 1) whether the relationship of Twitter sentiment and REIT returns is driven by fundamentals (rational sentiment), noise (irrational sentiment), or a combination of both, and 2) whether this relationship changed during the COVID period.

We find that the explanatory power of macroeconomic and capital market fundamentals such as inflation, stock market returns, credit risk spread, individual stock investor sentiment, and volatility is higher for optimistic, pessimistic, and overall Twitter sentiment in the COVID period than in the pre-COVID period. Furthermore, we provide evidence that pessimistic Twitter sentiment, as opposed to optimistic or overall sentiment, has the highest explanatory power for

REIT returns. Hereby, irrational pessimistic Twitter sentiment predicts REIT returns in the pre-COVID period, while rational (fundamentals-based) pessimistic Twitter sentiment predicts REIT returns in the COVID period.

Our study adds to the general finance and REIT literature in several ways. First, our study adds to a wider finance literature on social media and individual stock market investors (e.g., Chen et al., 2014; Cookson et al., 2024; Jame et al., 2016; Schmidt, 2020) by showing that the informative value of social media-based investor sentiment for stock returns can change over time, for example, due to individual investors becoming more informed about an industry. Our findings also emphasize the importance of separating sentiment measures into fundamental and non-fundamentals-based components.

Compared to institutional REIT investor sentiment (e.g., Aroul et al., 2023; Carstens & Freybote, 2019; Das et al., 2015), the sentiment of individual REIT investors has been neglected in the REIT literature. One challenge of investigating individual REIT investor sentiment is the difficulty of obtaining data. Considering the heavy use of social media by individual investors, our REIT-specific Twitter-based sentiment measures represent a proxy for the sentiment of this type of investor. Additionally, our study is the first to investigate factors that drive individual REIT investor sentiment. As Twitter sentiment has been found to impact REIT returns (Paulus et al., 2025), insights into whether it is due to fundamental information or irrationality have practical implications for the investment and asset pricing decision-making of sophisticated REIT investors.

Our study has a few limitations due to data availability and access: First, we only focus on Twitter as a social media platform. Compared to other platforms such as Seeking Alpha or StockTwits, Twitter messages are shorter, which impacts their informative content (Cookson et al., 2024). Future studies with access to data from other social media platforms may use our findings as a starting point and investigate the informative value of other social media sentiment for REITs. Another limitation lies in the dataset, as it does not allow us to identify the user behind each tweet and further investigate their characteristics. Last, we are not able to extend our dataset to the post-COVID period due to the drastic change in Twitter API accessibility with its conversion to the X platform in early 2023.

The remainder of our study is structured as follows. Next, we review the relevant literature to provide a theoretical framework. It is followed by the discussion of our data and methodology, as well as the presentation of our results and a conclusion.

3.3 Literature Review

3.3.1 Individual Investor Behavior

Several studies investigate individual (retail) stock market investors and show that the trading behavior of individual investors is characterized by overconfidence, herding, underperformance, and a focus on attention-grabbing stocks. Barber and Odean (2000) provide evidence that trading by individual investors reflects investor overconfidence and underperforms benchmarks. Barber and Odean (2002) investigate individual investor behavior after switching from phone-based to online trading. They show that after going online, investors traded more actively, speculatively, and less profitably, which can be explained by overconfidence due to self-attribution bias and illusions of control and knowledge.

Barber and Odean (2008) show that individual investors focus on buying attention-grabbing stocks, such as stocks featured in the media or stocks with extreme one-day returns. This tendency results from bounded rationality, lower sophistication, and limited attention of individual investors when deciding on which stocks to buy from a large selection of possible options. The selling behavior of individual investors, however, is less affected by attention-grabbing stocks. On the other hand, professional investors such as professional money managers are less likely to buy attention-grabbing stocks. Da et al. (2011) developed an attention measure based on Google searches to investigate the focus of individual investors on attention-grabbing stocks. They provide evidence that this measure captures individual investor attention, primarily of less sophisticated ones, and can predict stock price developments and IPO returns. Kumar (2009) shows that, in contrast to institutional investors, certain individual investors prefer lottery-type stocks, particularly in deteriorating economic conditions. Individuals investing in lottery-type stocks are relatively unsophisticated and more likely to spend funds on state lotteries.

Social media, mobile internet, and FinTech innovations further promote individual investor behaviors such as herding. Barber and Odean (2008) investigate attention-grabbing events that influence individual investor trades by focusing on news, trading volume, and previous stock returns. The authors acknowledge that chat rooms or word of mouth represent other ways in which stocks can grab the attention of individual investors. Barber et al. (2009) provide additional evidence for individual investor herding. Barber et al. (2022) show that Robinhood investors are more likely to engage in trading high-attention stocks than other individual investors, which can be explained by the low levels of sophistication of investors using Robinhood and its features, as well as herding. The trading behavior of Robinhood users can forecast abnormal returns.

However, Kelly and Tetlock (2013) and Kaniel et al. (2012) provide evidence that individual investor trades can also contain fundamental and private information. One explanation is that individual investors understand consumer preferences or have knowledge of a particular

company due to employee relationships or geographical proximity (Kelly & Tetlock, 2013). Another is that even if individual investors have imprecise information, the aggregation of their trades may yield valuable information (Kaniel et al., 2012). A third explanation is that individual investors learn over time and improve their abilities, while others acknowledge their insufficiencies and cease to actively trade (Seru et al., 2010).

3.3.2 Individual Investors and Social Media

Individual investors have a variety of information sources, such as news, friends and family, brokers, analysts, investment newsletters, and other publications. Social media represents an additional source of information, which has increased in importance for individual investors over the last years (Eliner & Kobilov, 2023). An emerging finance literature investigates the question of how valuable information from social media is: Does it just represent cheap talk and naïve investor reactions or contain decision-relevant information for stock market investors?

Schmidt (2020) shows that the sharing of trustworthy investment information on social media depends on the investment horizon of investors. Short-term investors with long-term information have an incentive to share correct information (reliable stock tips) while long-term investors have an incentive to disseminate incorrect or valueless information (cheap talk). Jame et al. (2016) focus on the use of social media as a source of information and investigate crowdsourced earnings forecasts from Estimize. They show that these forecasts have informative value for future earnings and capture the market's expectations. Hereby, a larger number of contributors improves the informative value. The authors provide evidence that social media and crowdsourced information represent an alternative source of information for the stock market. Sprenger et al. (2014) use Twitter messages to identify firm-specific good and bad events and provide evidence that they reflect relevant news events.

Cookson et al. (2024) show that differences in social media platforms (Twitter, StockTwits, and Seeking Alpha) in terms of length of messages and users can lead to variations in the relevance of social media-generated information. Furthermore, they find that the information (signal) in social media is distinct from traditional media. Cookson and Niessner (2020) find that the language on social media (StockTwits) is influenced by prominent investors and that information diffuses slowly from sophisticated to unsophisticated users over time. Focusing on Seeking Alpha, Chen et al. (2014) investigate how investor opinions on social media affect stock returns and earnings surprises. They show that negative words on social media contain value-relevant information beyond what is provided by financial analysts and provide evidence for the wisdom of crowds.

3.3.3 Twitter Sentiment and Stock Market Fundamentals

An emerging literature focuses on social media sentiment and its informative value for stock market fundamentals at the aggregated or disaggregated level. One set of studies investigates firm-level Twitter sentiment: Ranco et al. (2015) extract firm-specific Twitter sentiment and find that financial Twitter data has a relationship with abnormal stock returns, depending on Twitter volume. Tan and Tas (2021) provide evidence that firm-specific Twitter sentiment predicts trading volume for U.S. and international companies. The authors show that social media sentiment has informative value for stock returns, which is independent of news sentiment. They conclude that social media sentiment represents an additional source of information for developing trading strategies.

Sul et al. (2017) distinguish firm-specific Twitter sentiment into positive and negative and find that both types of sentiment have informative value for future stock returns. In particular, they show that a trading strategy based on Twitter sentiment can yield meaningful economic returns. C. Gu and Kurov (2020) show that Twitter sentiment has predictive value for stock returns, which can be partly explained by sentiment providing new information about analyst recommendations, price targets, and quarterly earnings. Karampatsas et al. (2023) find that firm-specific Twitter and StockTwits sentiment have a relationship with abnormal stock returns during negative earnings surprises, particularly for firms that are difficult to value and arbitrage. Ammari et al. (2023) investigate Twitter sentiment and stock market liquidity during the COVID period and find that the pandemic changed the relationship between sentiment and liquidity.

Other studies focus on Twitter sentiment in the context of individual industries and/or financial indices: Reboredo and Ugolini (2018) focus on Twitter sentiment in the context of renewable energy stocks and find no impact on returns, volatility, and trading volumes, which suggests that the wisdom of the crowd has no pricing effects for this industry. However, Twitter sentiment creates feedback effects for volatility and trading volume. The authors identify correlations between Twitter sentiment and financial indices. Zeitun et al. (2023) focus on multiple industries and find that changes in Twitter sentiment have a larger impact on the healthcare, communications, material, consumer discretionary, energy, staples, and information technology sectors than on other industries such as financial, real estate, industrial, and utilities.

Paulus et al. (2025) is the first study to investigate Twitter sentiment in the context of REITs. In line with previous non-REIT studies, they show that Twitter sentiment predicts aggregated REIT returns. However, the authors fail to investigate why Twitter sentiment has explanatory power for REIT returns. Twitter sentiment may contain fundamental information, which in turn explains its informative value for REIT returns. It may also contain noise (irrationality), which affects REIT

returns. The relationship of Twitter sentiment and REIT returns could also be driven by a combination of fundamental information and noise.

The finance literature on individual investors is divided on whether this type of investor represents unsophisticated noise traders prone to behavioral biases (e.g., Barber et al., 2009; Barber & Odean, 2000, 2002) or investors trading on fundamental information that learn by trading (Kaniel et al., 2012; Kelly & Tetlock, 2013; Seru et al., 2010). We firstly hypothesize that fundamentals have a higher explanatory power for REIT-specific Twitter sentiment in the COVID period as opposed to the pre-pandemic period. During the pandemic, commercial real estate received a lot of attention from the media, which likely increased the attention REITs received from individual investors. In times of lockdowns and COVID financial support, these individual investors had more time to research stocks and increase their understanding of the REIT industry and its fundamentals. As a result, we also expect the rational (fundamentals-based) component of Twitter sentiment to have a higher explanatory power than irrational Twitter sentiment for REIT returns in the COVID period. On the other hand, considering that individual users, as proxied by Twitter users, are expected to be less familiar with REITs in the pre-pandemic period, we hypothesize that the relationship between Twitter sentiment and REIT returns was driven by irrationality in this period.

3.4 Data and Methodology

3.4.1 Derivation of Twitter Sentiment Measures

To derive our Twitter sentiment measures, we collected textual data using the Tweepy API for Python with the firm identifiers or names of U.S. REITs as search parameters over the period of January 2014 to December 2022. The initial dataset consists of about 3.9 million tweets before data cleaning. As Twitter messages can be categorized as informal language, it is necessary to apply extensive natural language processing techniques in order to derive their correct sentiment (e.g., Ankit & Saleena, 2018; Dorle & Pise, 2018).

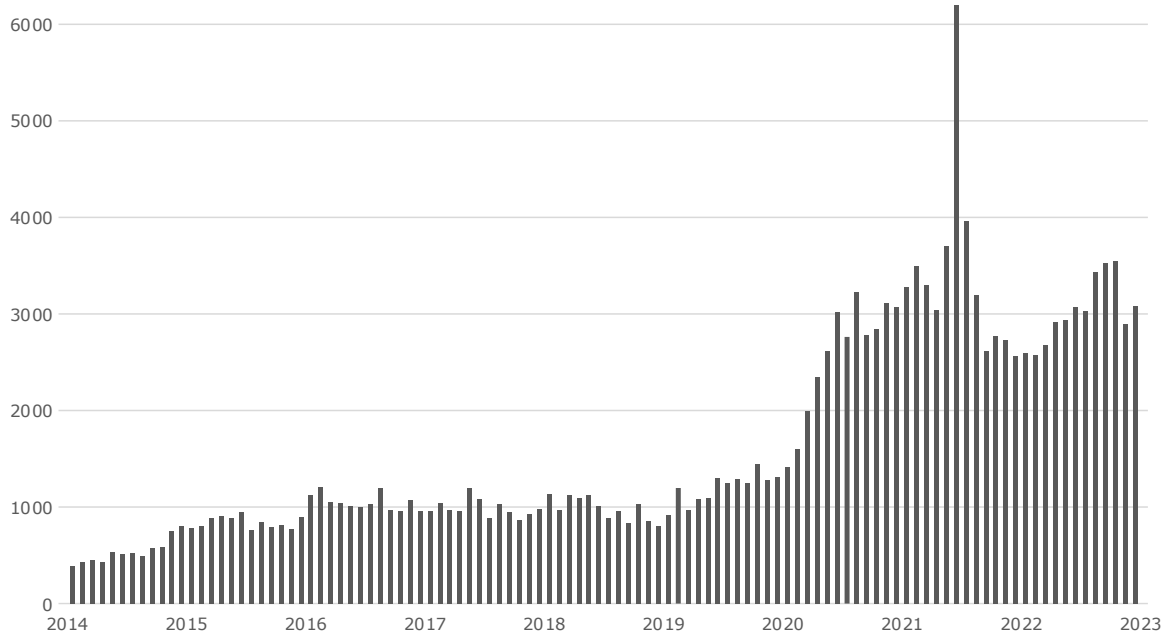
We follow the approach employed by Paulus et al. (2025) to preprocess and filter the Twitter data: We convert all words to lowercase, followed by the conversion of slang terms to their formal counterparts by utilizing the Internet & Slang Dictionary from www.noslang.com, consisting of over 5,000 slang terms (Jones, 2022). Further, we expand contractions such as “isn’t” to “is not”. To address the unique characteristics of the language used on Twitter, mentions signalized by “@word”, cashtags “$word” and hashtags “#word”, retweets, links/URLs, emojis, and non-ASCII symbols are removed. Additionally, numbers and punctuation are taken out of the textual corpus. The final preprocessing step applied is lemmatization to

transform the words into their root forms, reducing the complexity and size of the vocabulary of the textual data (Manning et al., 2008; Paulus et al., 2025).

In the derivation of our Twitter sentiment measures, we place particular emphasis on the selection process to minimize irrelevant messages in the textual corpus. After excluding simple retweets and messages without interactions, duplicate messages are eliminated in two steps: First, messages that are identical are filtered out. Second, messages that are mostly identical, indicating automated spam during the same day, are removed based on their word sequences. The temporal restriction minimizes the risk of removing relevant tweets that coincidentally share the same word sequence. Last, we use the Global Vector (GloVe) and cosine similarity approach for further filtering the textual data to increase its relevance to the REIT market (Paulus et al., 2025). Following this approach, we create and combine multiple lists of signal words related to REIT-related phrases to achieve a more extensive and simultaneously relevant list compared to Paulus et al. (2025). As the unfiltered Twitter data contains messages related to cryptocurrency or other digital asset-related content, we repeat this process and create word lists related to these topics. By only including tweets that contain at least one word with a REIT context and no phrase from the other word lists, we avoid the inclusion of leftover irrelevant messages. The final dataset contains 175,130 remaining messages. As shown in Figure 3.1, the monthly REIT-specific Twitter message volume sharply increased with the start of the COVID-19 pandemic. This provides support for our argument that REITs received increased attention from individual investors during the pandemic period, as suggested by increased activity with regard to REIT-specific tweets.

We derive our Twitter sentiment measures based on the cleaned Twitter data using the Robustly Optimized BERT Pretraining Approach (RoBERTa) model by Liu et al. (2019), which is an improved version of the Bidirectional Encoder Representations from Transformers (BERT), introduced by Devlin et al. (2019). These models are large language models (LLMs) and represent one of the state-of-the-art approaches in natural language processing (Karanikola et al., 2023). It is noteworthy that the GPT model which the industry-disrupting ChatGPT is based on (Brown et al., 2020) falls under the LLM approach. Additionally, the usage of the RoBERTa model minimizes the necessity to obtain an adequate training dataset as the model comes pre-trained and is only optimized in relation to an individual use case (Liu et al., 2019). Considering that Paulus et al. (2025) find more sophisticated classifiers to yield better performance for sentiment analysis, we expect the RoBERTa model to provide an even more accurate representation of the true sentiment embedded in Twitter messages.

Figure 3.1 – REIT-Specific Twitter Message Volume per Month



Note: The figure shows the monthly REIT-specific Twitter Message Volume over the period from January 2014 to December 2022 after applying our filter process. The y-axis represents the absolute monthly number of tweets, while the x-axis denotes the timeline in years. A steady upward trend is observable from 2014 through 2019, followed by a pronounced increase beginning in early 2020. After the peak in early 2021, the values stabilized at an elevated level with continued fluctuations.

Our approach to using RoBERTa for sentiment classification results in a probabilistic output with values between 0 and 1, where values close to 0 indicate a low probability of a message being positive and vice versa. To achieve a three-way classification, we deploy a decision rule to achieve a three-level sentiment classification in line with Paulus et al. (2025), given as:

$$P \geq 0.667 \rightarrow \text{positive sentiment}$$

$$0.333 < P < 0.667 \rightarrow \text{neutral sentiment}$$

$$P \leq 0.333 \rightarrow \text{negative sentiment}$$

After classifying the Twitter data, we aggregate it on a monthly frequency to construct three Twitter sentiment measures in line with previous studies (e.g., Hausler et al., 2018; Paulus et al., 2025; Ruscheinsky et al., 2018). The first measure, the Optimism Indicator (*OI*), is based on Antweiler and Frank (2004) and captures the overall optimism expressed in the messages. The *OI* is defined as shown in Equation 3.1.

$$OI_t = \frac{\sum_{i=1}^{I} \text{positive message}_{i,t}}{\text{total number of messages}_t} \tag{3.1}$$

Acting as the reciprocal of the *OI*, the second measure, the Pessimism Indicator (*PI*), measures the total amount of negative messages in relation to the total number of messages as shown in Equation 3.2.

It is based on the idea of a negativity bias as proposed by Rozin and Royzman (2001), who stated an increased effect of negative influences in comparison to positive ones.

$$PI_t = \frac{\sum_{j=1}^{J} \text{negative message}_{j,t}}{\text{total number of messages}_t} \tag{3.2}$$

The third sentiment measure derived is the Sentiment Quotient (*SQ*), which measures the ratio of positive messages to the sum of positive and negative messages, as shown in Equation 3.3.

$$SQ_t = \frac{\sum_{i=1}^{I} \text{positive message}_{i,t}}{\sum_{i=1}^{I} \text{positive message}_{i,t} + \sum_{j=1}^{J} \text{negative message}_{j,t}} \tag{3.3}$$

For all three measures, i represents a positive classified message, j a negative message and t the aggregated time frame.

3.4.2 Macroeconomic, Capital and Real Estate Market Variables

We obtain the following macroeconomic, capital, and commercial real estate market variables over the period of 2014 to 2022. These variables are on a monthly basis to balance the short-term and quick-reacting nature of the stock market and the slower-changing underlying real estate market. Furthermore, this ensures capturing the true change in value of the underlying companies as these can only be seen in the long term, as the short-term speculative asset prices are almost random walks on a day-to-day basis (Shiller, 2017).

To capture macroeconomic conditions, we obtain the monthly U.S. inflation (*Inflation*) as well as the unemployment rate (*Unrate*) from the U.S. Bureau of Labor Statistics. Furthermore, stock and bond market conditions may drive Twitter sentiment. Consequently, we obtain the monthly return on the S&P 500 Index (*SP500*) to capture stock market conditions and the monthly spread between Moody's Seasoned Baa- and Aaa-rated corporate bonds (*Spread*) to capture bond market conditions from S&P Global and the Federal Reserve Bank of St. Louis. Considering that Twitter sentiment reflects individual investor sentiment, it may also be affected by general stock market sentiment, as proxied by the monthly American Association of Individual Investors (*AAII*) bearish sentiment, or the volatility in the stock market, as proxied by the monthly CBOE Volatility Index (*VIX*). Last, the sentiment of Twitter users about REITs may be influenced by the performance of commercial real estate markets. We capture these conditions with the U.S. composite of the CoStar Commercial Repeat Sales Index (*CCRSI*). As our dependent variable, and similar to Paulus et al. (2025), we use the total return of a broad FTSE Nareit index, specifically, the FTSE Nareit All Equity REITs Index, for our analysis of the relationships of rational and irrational Twitter sentiment with REIT returns. Our study hereby differs from Paulus et al. (2025), who use the total return on the FTSE Nareit All REITs Index, which also includes mortgage and hybrid REITs.

We test for the presence of a unit root in our time series data using the Augmented Dickey-Fuller Test (ADF). To ensure the required stationarity, uniformity, and ease of interpretation, all variables are considered as first differences or growth rates. Additionally, we conduct tests for serial correlation and heteroscedasticity and apply Newey and West (1987) standard errors to increase robustness against those and statistical validity. Table 3.1 presents the descriptive statistics for our sentiment measures and fundamental variables.

Table 3.1 – Descriptive Statistics

Panel A: Pre-COVID Period					
	Min	**Max**	**Mean**	**Median**	**Std. Dev.**
Total Return	-0.079	0.116	0.010	0.010	0.037
OI	0.520	0.692	0.586	0.582	0.037
PI	0.107	0.237	0.181	0.183	0.026
SQ	0.687	0.866	0.764	0.765	0.035
Unrate	0.035	0.067	0.047	0.047	0.009
Inflation	-0.002	0.029	0.016	0.017	0.008
SP500	-0.092	0.083	0.008	0.010	0.033
Spread	0.550	1.490	0.901	0.890	0.220
AAII	0.197	0.442	0.294	0.283	0.049
VIX	10.125	24.953	14.979	14.241	3.437
CCRSI	141.893	230.276	188.060	186.404	27.675
Panel B: COVID Period					
	Min	**Max**	**Mean**	**Median**	**Std. Dev.**
Total Return	-0.187	0.096	0.002	0.017	0.066
OI	0.399	0.576	0.524	0.535	0.036
PI	0.169	0.319	0.217	0.211	0.029
SQ	0.556	0.768	0.707	0.714	0.042
Unrate	0.035	0.148	0.058	0.047	0.028
Inflation	0.001	0.091	0.047	0.054	0.031
SP500	-0.125	0.127	0.007	0.022	0.062
Spread	0.650	1.700	0.940	0.900	0.260
AAII	0.225	0.543	0.383	0.400	0.096
VIX	16.957	57.737	25.117	23.299	7.633
CCRSI	231.056	307.702	268.269	265.382	27.822

Note: The table presents the descriptive statistics for the pre-COVID period (01/2014 - 01/2020) and the COVID period (02/2020 – 12/2022). *Total Return* represents the monthly total return of the FTSE Nareit All Equity REITs Index. *OI* is the monthly Twitter Optimism Indicator, *PI* is the Pessimism Indicator, and *SQ* is the Sentiment Quotient that reflects the general sentiment. *Inflation* is the monthly inflation, and *Unrate* is the monthly unemployment rate in the U.S. *SP500* represents the monthly return of the general stock market as the monthly return on the S&P 500 Index. *Spread* is the monthly spread of Moody's Seasoned Baa- and Aaa-rated corporate bonds, *AAII* is the monthly general stock market sentiment measured by the American Association of Individual Investors, and *VIX* is the stock market volatility proxied by the CBOE Volatility Index. *CCRSI* is the monthly CoStar Commercial Repeat Sales Index. The values are shown as level data.

Considering we expect a structural break in the relationship between Twitter sentiment and REIT returns as a result of the COVID-19 pandemic, we separate our sample into two periods: The pre-COVID period is defined as January 2014 to January 2020, and the COVID period covers February 2020 to December 2022. The longer time horizon of the pre-crisis period enables a stabilized market overview under normal conditions.

3.4.3 Estimation

Our empirical investigation has two components: First, we assess the relationships of fundamental variables with optimistic, pessimistic, and overall Twitter sentiment to determine

their informative value for Twitter sentiment. Second, we separate sentiment measures into rational and irrational components and investigate their relationships with aggregated REIT returns.

For the first part, we use the sentiment measures, *OI*, *PI*, and *SQ* as dependent variables, and fundamental variables as independent variables, separated by the pre-COVID and COVID periods. As we expect an immediate impact of Twitter sentiment due to the platform's quick-moving nature, we assume a non-lagged relationship between fundamental variables and Twitter sentiment. To achieve the needed flexibility to account for such a non-lagged relationship as well as to address endogeneity, we employ an autoregressive distributed lag (ARDL) model as shown in Equation 3.4. To simplify the model selection process, we assume the lag order of the autoregressive component and the independent regressor to be constant.

$$y_t = \alpha + \sum_{i=1}^{l} \phi_i \, y_{t-i} + \sum_{i=0}^{l} \beta_i \, x_{t-i} + u_t \tag{3.4}$$

with y_t resembling one of the three sentiment measures (*OI*, *PI*, *SQ*) and x_t as one of the respective fundamental variables (*Unrate*, *Inflation*, *SP500*, *Spread*, *AAII*, *VIX*, *CCRSI*).

We employ statistical tests such as the Akaike Information Criterion (AIC), Bayesian Information Criterion (BIC), and Final Prediction Error (FPE) to select an appropriate lag order. To ensure comparability, the lag order within the stated sets of regressions is kept consistent. Therefore, we determined the optimal lag order for each regression of the sets individually and compared them. The optimal lag order for the regressions within these sets was mostly unambiguous. If it was not, we utilized a majority voting system for each set.

For the second part of our empirical analysis, we extract the rational and irrational components of the Twitter sentiment using an orthogonalization approach based on Ling et al. (2014). In particular, we first regress the Twitter sentiment measures on our fundamental variables (*Unrate*, *Inflation*, *SP500*, *Spread*, *AAII*, *VIX*, *CCRSI*) and retain the fitted values. This represents the rational component of Twitter sentiment (*OI*, *PI*, *SQ*) that can be explained by macroeconomic, capital market, and commercial real estate market conditions. The residuals from these regressions represent the irrational components of Twitter sentiment as they are not explained by fundamentals but rather other factors such as irrationality.

Considering that the impact of rational and irrational Twitter sentiment on REIT returns is not likely to be immediate, it is not necessary to include non-lagged values for this analysis. Therefore, we employ a VARX model as shown in Equation 3.5. The VARX represents an extended form of the regular vector autoregression (VAR) framework, suitable when control variables are included. We include a set of control variables in our time series analysis in line with Brooks and Tsolacos (2010) to increase the robustness of our results. In particular, we control for macroeconomic

conditions that impact REIT returns, such as *Inflation* and *Unrate*. We also include the U.S. Economic Policy Uncertainty Index (*Uncertainty*) (Baker et al., 2024) as policy uncertainty, especially during a crisis, can impact commercial real estate markets and thus REIT returns. Furthermore, we control for debt capital market conditions impacting REIT market conditions by including *Spread* and the difference between the 10-year Treasury bond yield and the 3-month Treasury bill (*Term*) in line with previous studies (e.g., Braun et al., 2020; Clayton et al., 2009; Freybote & Seagraves, 2017; Hausler et al., 2018; Hoesli et al., 2008; Lin et al., 2009; Paulus et al., 2025).

$$y_t = A_0 + \sum_{i=1}^{l} A_i\, y_{t-i} + BZ_t + u_t \tag{3.5}$$

where Z_t is the vector of exogeneous variables.

The lag order selection process for the VARX models follows the process described previously. To ensure further comparability, the lag order is determined for the regression using the aggregated sentiment before orthogonalizing and then also applied to the derived components. Additionally, we introduce a minimum lag order of six months to ensure capturing potential long-term relationships, as found by Paulus et al. (2025). Lastly, we use Granger causality tests, which test the joint significance of two lagged endogenous variables (Granger, 1969), to assess whether one variable in our VAR model predicts the other variable, suggesting a cause-and-effect relationship between the two (Jalil & Rao, 2019).

3.5 Results

3.5.1 Twitter Sentiment and Fundamental Variables

Table 3.2 presents the results for inflation and unemployment based on the model in Equation 3.4, separated by the pre-COVID and COVID periods. In the pre-COVID period, inflation had no explanatory power for REIT-specific Twitter sentiment, irrespective of the measure used (*OI*, *PI*, *SQ*). In the COVID period, inflation predicts optimistic, pessimistic, and overall Twitter sentiment in multiple lags. However, the directionality of this relationship is ambiguous as signs vary. In particular, inflation has a negative and positive relationship with optimistic, pessimistic, and overall REIT-specific Twitter sentiment, respectively, in different lags. This pattern shows an immediate reaction with changing signs in lagged values. This reverse effect could be interpreted as an overreaction first and a correction movement later. This dynamic could also be attributed to an increased awareness of individuals and a more pronounced impact of inflation during the COVID period. In the pre-COVID period, inflation in the U.S. was stable and moderate, whereas inflation rates increased dramatically during the pandemic due to supply chain disruptions, increased consumer demand, and higher energy prices.

The unemployment rate has limited informative value for optimistic, pessimistic, and overall Twitter sentiment in either period. It is noteworthy that the direction of significant coefficients flips between the pre-COVID and COVID period. For example, unemployment has a significantly negative coefficient (first lag), albeit only at the 10% level, for *SQ* in the pre-COVID period, but a significantly positive coefficient (second lag) in the COVID period. The results for stock and bond market conditions, as proxied by the S&P 500 return and credit risk spread, respectively, are presented in Table 3.3. The return on the S&P 500 Index has a negative (positive) relationship with optimistic and overall (pessimistic) Twitter sentiment in its first lag in the pre-COVID period. Thus, an increase in the S&P index return in the previous month predicts a decrease in the optimistic and overall REIT-specific sentiment of Twitter users and an increase in the pessimistic one. One explanation for this finding is that individual investors in this period considered REITs an alternative asset class to invest in when index funds and other stocks performed poorly. On the other hand, our results for SP500 in the COVID period are in line with expectations: An increase in the S&P 500 index return in the previous and current month predicts an increase (decrease) in the optimistic and overall (pessimistic) REIT-specific Twitter sentiment.

Table 3.2 – ARDL Results: Sentiment, Inflation, and Unemployment Rate

Panel A: Inflation						
	Optimism Indicator (OI)		**Pessimism Indicator (PI)**		**Sentiment Quotient (SQ)**	
	Pre-COVID	**COVID**	**Pre-COVID**	**COVID**	**Pre-COVID**	**COVID**
$Sentiment_{t-1}$	-0.633***	-0.634***	-0.451***	-0.668***	-0.495***	-0.630***
$Sentiment_{t-2}$	-0.370***	-0.163*	-0.360***	-0.164**	-0.370***	-0.145*
$Sentiment_{t-3}$		0.005		0.037		0.020
$Inflation_{t-0}$	-0.976	1.335**	1.081	-2.023***	-1.440	2.486***
$Inflation_{t-1}$	-1.187	-0.894	0.118	1.032***	-0.446	-1.449***
$Inflation_{t-2}$	1.932	-0.474	-1.383	0.401	1.947	-0.563
$Inflation_{t-3}$		-2.553***		2.216***		-3.091***
Constant	-0.003	0.009**	0.001	-0.005**	-0.002	0.009**
Adj. R^2	0.31	0.32	0.21	0.43	0.24	0.40
Panel B: Unemployment Rate						
	Optimism Indicator (OI)		**Pessimism Indicator (PI)**		**Sentiment Quotient (SQ)**	
	Pre-COVID	**COVID**	**Pre-COVID**	**COVID**	**Pre-COVID**	**COVID**
$Sentiment_{t-1}$	-0.616***	-0.445**	-0.474***	-0.247	-0.516***	-0.268
$Sentiment_{t-2}$	-0.398***	-0.102	-0.348***	-0.119	-0.369***	-0.060
$Unrate_{t-0}$	-2.090	-0.301	1.224	-0.163	-1.811	0.107
$Unrate_{t-1}$	-0.372	-0.002	4.865**	0.077	-5.102*	0.015
$Unrate_{t-2}$	-2.185	0.314***	2.603	-0.399***	-3.243	0.503***
Constant	-0.005*	0.003	0.005*	-0.002	-0.007*	0.003
Adj. R^2	0.29	0.11	0.23	0.20	0.24	0.16

Note: This table shows the results of autoregressive distributed lag models (ARDL). In Panel A, the market fundamental variable inflation (*Inflation*) is regressed on the different sentiment measures *OI*, *PI*, and *SQ* as the dependent variables for the pre-COVID and COVID period. Panel B considers the unemployment rate (*Unrate*), respectively. ***, **, and * denote significance at the 1%, 5% and 10% level.

Corporate bond market conditions, as proxied by the credit risk spread, have no informative value for REIT-specific Twitter sentiment in the pre-COVID period. However, they predict Twitter sentiment in the COVID period in the expected directions: An increase in the credit risk spread in

the current and previous month predicts a decrease (increase) in the REIT-specific optimistic and overall (pessimistic) Twitter sentiment. An explanation for this finding is that during the COVID period, media reporting focused on tenant and landlord defaults to a larger degree, which in turn improved the understanding of individual investors as it relates to credit risk and REITs.

Table 3.3 – ARDL Results: Sentiment, S&P 500 Returns and Credit Risk Spread

Panel A: S&P 500 Returns						
	Optimism Indicator (OI)		**Pessimism Indicator (PI)**		**Sentiment Quotient (SQ)**	
	Pre-COVID	**COVID**	**Pre-COVID**	**COVID**	**Pre-COVID**	**COVID**
$Sentiment_{t-1}$	-0.680***	-0.411***	-0.457***	-0.419***	-0.515***	-0.385***
$Sentiment_{t-2}$	-0.409***	-0.067	-0.394***	-0.030	-0.411***	-0.024
$SP500_{t-0}$	-0.126	0.165***	0.033	-0.167***	-0.072	0.225***
$SP500_{t-1}$	-0.304***	0.178**	0.138*	-0.130**	-0.226***	0.188**
$SP500_{t-2}$	-0.031	0.051	-0.031	-0.018	0.022	0.031
Constant	0.001	-0.001	0.000	0.002	0.000	-0.002
Adj. R^2	0.36	0.23	0.21	0.33	0.25	0.31
Panel B: Credit Risk Spread						
	Optimism Indicator (OI)		**Pessimism Indicator (PI)**		**Sentiment Quotient (SQ)**	
	Pre-COVID	**COVID**	**Pre-COVID**	**COVID**	**Pre-COVID**	**COVID**
$Sentiment_{t-1}$	-0.642***	-0.623***	-0.427***	-0.627***	-0.474***	-0.602***
$Sentiment_{t-2}$	-0.381***	-0.250***	-0.330***	-0.228*	-0.344***	-0.209*
$Spread_{t-0}$	0.009	-0.072*	-0.065	0.056**	0.075	-0.082**
$Spread_{t-1}$	-0.008	-0.075***	0.063	0.046**	-0.073	-0.069**
$Spread_{t-2}$	-0.064	0.022	0.018	-0.020	-0.035	0.032
Constant	-0.003	0.003	0.001	-0.002	-0.002	0.003
Adj. R^2	0.29	0.22	0.20	0.20	0.23	0.21

Note: This table shows the results of autoregressive distributed lag models (ARDL). In Panel A, the market fundamental variable S&P 500 returns (*SP500*) is regressed on the different sentiment measures *OI*, *PI*, and *SQ* as the dependent variables for the pre-COVID and COVID period. Panel B considers the credit risk spread (*Spread*), respectively. ***, **, and * denote significance at the 1%, 5% and 10% level.

As shown in Table 3.4, the bearish individual investor sentiment (*AAII*) and volatility index (*VIX*) have no informative value for REIT-specific Twitter sentiment in the pre-COVID period, with the exception of *AAII* and *OI*. In the COVID period, the general individual stock investor sentiment (*AAII*) predicts optimistic, pessimistic, and overall Twitter sentiment. However, the directionality varies for *OI* and *SQ*. *AAII* bearish sentiment has a significantly negative coefficient for *OI* and *SQ* in the current or previous month and a significantly positive one for *PI* in the current month. Thus, an increase in bearish individual stock investor sentiment in the current or previous month predicts a decrease in the optimistic and overall Twitter sentiment and an increase in the pessimistic one. It is noteworthy that the second lag for *AAII* in the *OI* and *SQ* regressions has a significantly positive coefficient, albeit only at the 10% level.

Our findings suggest that, to a certain degree, changes in individual stock investor sentiment, as captured by *AAII*, are mirrored by the REIT-specific sentiment of individual investors observed on Twitter. However, the explanatory power of *AAII* sentiment for the Twitter-based sentiment is relatively low. One explanation is that the *AAII* sentiment captures general stock market sentiment while our Twitter sentiment captures REIT-specific sentiment of individual investors.

Another explanation is the characteristics of individual investors responding to the *AAII* survey, which may be different from the characteristics of individual investors using Twitter (e.g., sophistication, experience, access to other data).

Table 3.4 – ARDL Results: Sentiment, Stock Market Sentiment, and Stock Market Volatility

	Panel A: AAII Bearish					
	Optimism Indicator (OI)		Pessimism Indicator (PI)		Sentiment Quotient (SQ)	
	Pre-COVID	COVID	Pre-COVID	COVID	Pre-COVID	COVID
$Sentiment_{t-1}$	-0.647***	-0.309**	-0.445***	-0.297**	-0.500***	-0.277***
$Sentiment_{t-2}$	-0.402***	0.056	-0.370***	0.061	-0.386***	0.079
$AAII_{t-0}$	0.012	-0.076	-0.061	0.089***	0.062	-0.109**
$AAII_{t-1}$	0.061	-0.066*	-0.019	0.007	0.041	-0.029
$AAII_{t-2}$	0.117**	0.107*	-0.065	-0.070	0.097	0.118*
Constant	-0.003	0.003	0.002	-0.002	-0.003	0.003
Adj. R^2	0.31	0.18	0.20	0.22	0.23	0.20
	Panel B: VIX					
	Optimism Indicator (OI)		Pessimism Indicator (PI)		Sentiment Quotient (SQ)	
	Pre-COVID	COVID	Pre-COVID	COVID	Pre-COVID	COVID
$Sentiment_{t-1}$	-0.645***	-0.461***	-0.447***	-0.399***	-0.498***	-0.401***
$Sentiment_{t-2}$	-0.406***		-0.370***		-0.388***	
VIX_{t-0}	-0.001	-0.003***	0.000	0.003***	0.000	-0.004***
VIX_{t-1}	0.001	-0.002***	0.000	0.001**	0.000	-0.001***
VIX_{t-2}	0.001		-0.001		0.001	
Constant	-0.003	0.000	0.002	0.001	-0.002	-0.001
Adj. R^2	0.29	0.61	0.18	0.59	0.21	0.63

Note: This table shows the results of autoregressive distributed lag models (ARDL). In Panel A, the market fundamental variable stock market sentiment (*AAII*) is regressed on the different sentiment measures *OI*, *PI*, and *SQ* as the dependent variables for the pre-COVID and COVID period. Panel B considers the VIX index (*VIX*), respectively. ***, **, and * denote significance at the 1%, 5% and 10% level.

During COVID, the volatility index has a negative relationship with the optimistic and overall Twitter sentiment and a positive one with the pessimistic sentiment, which is in line with expectations. An increase in volatility in the stock market in the current and previous months predicts a decrease (increase) in the optimistic and overall (pessimistic) sentiment of Twitter users about REITs.

Last, we present the results for commercial real estate market conditions as proxied by the *CCRSI* index in Table 3.5. They suggest limited informative value of commercial real estate market conditions for Twitter sentiment in the pre-COVID and COVID periods. While one coefficient is significant for each sentiment measure and period, with two being significant for *OI* in the COVID period, several are only significant at the 10% level. Furthermore, the results suggest that commercial real estate market conditions and Twitter sentiment have an inverse relationship, irrespective of period: An increase in commercial real estate prices in the current and/or previous two months predicts a decrease in the optimistic and overall REIT-specific Twitter sentiment and an increase in the pessimistic one.

Table 3.5 – ARDL Results: Sentiment and Commercial Real Estate Prices

	CCRSI					
	Optimism Indicator (OI)		Pessimism Indicator (PI)		Sentiment Quotient (SQ)	
	Pre-COVID	COVID	Pre-COVID	COVID	Pre-COVID	COVID
$Sentiment_{t-1}$	-0.628***	-0.376***	-0.472***	-0.429***	-0.506***	-0.386***
$Sentiment_{t-2}$	-0.404***	-0.099	-0.425***	-0.072	-0.431***	-0.066
$CCRSI_{t-0}$	0.003	-0.003*	-0.004	0.002	0.005	-0.003
$CCRSI_{t-1}$	-0.006*	0.001	0.008**	0.000	-0.010**	0.001
$CCRSI_{t-2}$	0.004	-0.004***	-0.003	0.003*	0.004	-0.004**
Constant	-0.005	0.017***	0.000	-0.011*	-0.001	0.017**
Adj. R^2	0.30	0.28	0.27	0.22	0.28	0.25

Note: This table shows the results of autoregressive distributed lag models (ARDL). The market fundamental variable Commercial Real Estate Prices (*CCRSI*) is regressed on the different sentiment measures, *OI*, *PI*, and *SQ*, as the dependent variables for the pre-COVID and COVID period. ***, **, and * denote significance at the 1%, 5% and 10% level.

This may suggest that individual investors lack the sophistication with regard to commercial real estate markets to understand the effect of asset prices on REIT returns and profitability. Another possible explanation is that these investors consider high property prices in the direct real estate market an issue for the growth opportunities of REITs. While an investigation into the sophistication of individual REIT investors with regard to commercial real estate markets is beyond the scope of this study, future studies may use our findings as a starting point to further investigate these and other explanations.

Overall, our results in Tables 3.2 to 3.5 suggest that fundamental variables have limited explanatory power for optimistic, pessimistic, and overall Twitter sentiment in the pre-COVID period. The informative value of fundamentals such as inflation, stock market returns, individual stock investor sentiment, volatility, and credit risk spread for REIT-specific Twitter sentiment increased in the COVID period, which is in line with our expectations.

3.5.2 The Relationships of Rational and Irrational Twitter Sentiment and REIT Returns

Next, we investigate whether the relationship between Twitter sentiment and REIT returns in the pre-COVID and COVID period is driven by fundamentals (rational sentiment component), irrationality (irrational component), or both. We present our results for the respective rational and irrational components of the respective sentiment measure (*OI*, *PI*, *SQ*) in the pre-COVID period in Table 3.6.

They suggest that primarily irrational pessimistic Twitter sentiment has explanatory power for REIT returns in this period. While the coefficient on the second lag of irrational *SQ* is significant, it is only at the 10% level. Furthermore, the coefficient on the fifth lag of rational *OI* is significant. Both results are likely of negligible substantial significance.

Table 3.6 – VAR Results: Rational and Irrational Sentiment and REIT Returns (Pre-COVID)

	FTSE Nareit All Equity REITs Total Return					
	Optimism Indicator (OI)		Pessimism Indicator (PI)		Sentiment Quotient (SQ)	
	Rational	**Irrational**	**Rational**	**Irrational**	**Rational**	**Irrational**
Total Return$_{t-1}$	-0.293**	-0.331	-0.328**	-0.354*	-0.344***	-0.351
Total Return$_{t-2}$	-0.021	-0.036	0.014	-0.030	0.002	-0.036
Total Return$_{t-3}$	-0.161**	-0.195**	-0.135	-0.183**	-0.119	-0.195**
Total Return$_{t-4}$	-0.269***	-0.342**	-0.303***	-0.333**	-0.319***	-0.345**
Total Return$_{t-5}$	-0.146	-0.103	-0.169	-0.121	-0.138	-0.119
Total Return$_{t-6}$	-0.049	-0.057	-0.062	-0.039	-0.070	-0.048
Sentiment$_{t-1}$	0.396	-0.168	0.294	0.241	0.055	-0.184
Sentiment$_{t-2}$	0.233	-0.108	-0.772	0.355**	0.605	-0.237*
Sentiment$_{t-3}$	-0.144	0.019	-0.139	0.247	0.080	-0.140
Sentiment$_{t-4}$	0.119	-0.091	0.868	0.340*	-0.374	-0.206
Sentiment$_{t-5}$	-0.671**	-0.154	0.619	0.276	-0.819	-0.195
Sentiment$_{t-6}$	0.537	0.063	0.249	0.241	0.014	-0.105
Constant	0.018***	0.019**	0.018***	0.019**	0.018***	0.019**
Controls	*YES*	*YES*	*YES*	*YES*	*YES*	*YES*
Adj. R²	0.15	0.09	0.11	0.09	0.10	0.08
	Granger Causality					
Sentiment	0.31	0.76	0.63	0.78	0.65	0.84
Total Return	0.21	0.14	0.88	0.84	0.92	0.71

Note: This table shows the results for the vector autoregression models with exogenous variables (VARX). The models are tested for each sentiment measure (*OI*, *PI*, *SQ*) in the pre-COVID period (01/2014 - 01/2020). The sentiment has been separated into rational and irrational components using an orthogonalization approach based on Ling et al. (2014). The total return of the FTSE Nareit All Equity REITs Index and the sentiment measure each show the lagged results for the past six months. To control for external effects, the control variables *Inflation*, *Unrate*, *Uncertainty*, *Spread*, and *Term* are integrated in the models. Besides the VAR models, the Granger Causality between the independent variable and each sentiment measure is tested, and the respective p-values are presented. ***, **, and * denote significance at the 1%, 5% and 10% level.

On the other hand, rational pessimistic REIT sentiment has explanatory power for aggregated REIT returns in the COVID period (Table 3.7). An increase in the rational pessimistic Twitter sentiment in the previous two months predicts a decrease in REIT returns. While the coefficient on the second lag of rational *SQ* is significant, it is only at the 10% level.

Table 3.7 – VAR Results: Rational and Irrational Sentiment and REIT Returns (COVID)

	FTSE Nareit All Equity REITs Total Return					
	Optimism Indicator (OI)		Pessimism Indicator (PI)		Sentiment Quotient (SQ)	
	Rational	**Irrational**	**Rational**	**Irrational**	**Rational**	**Irrational**
Total Return$_{t-1}$	-0.383**	-0.378***	-0.842***	-0.289*	-0.617**	-0.293*
Total Return$_{t-2}$	-0.179	-0.049	-0.516**	0.286	-0.352	0.196
Total Return$_{t-3}$	0.133	0.240	0.105	0.381*	0.124	0.318
Total Return$_{t-4}$	0.632***	0.700***	0.723***	0.916***	0.677***	0.801***
Total Return$_{t-5}$	0.144	0.345***	0.559	0.232**	0.330	0.241**
Total Return$_{t-6}$	-0.657***	-0.290**	-0.428	-0.539***	-0.553*	-0.477***
Sentiment$_{t-1}$	-0.237	0.569	-1.553**	-2.103*	0.528	1.174
Sentiment$_{t-2}$	1.000	0.188	-1.522**	-0.526	0.977*	0.336
Sentiment$_{t-3}$	-0.108	-0.180	-0.432	0.652***	0.080	-0.444*
Sentiment$_{t-4}$	0.141	0.556	0.226	-0.569	-0.059	0.394
Sentiment$_{t-5}$	0.020	-0.090	0.523	-0.077	-0.127	0.029
Sentiment$_{t-6}$	1.311***	-0.038	-1.336**	0.194	1.021**	-0.133
Constant	0.007	0.004	0.006	0.015**	0.007	0.012*
Controls	*YES*	*YES*	*YES*	*YES*	*YES*	*YES*
Adj. R²	0.45	0.33	0.54	0.42	0.51	0.40

To be continued

Table 3.7 continued

	Optimism Indicator (OI)		Pessimism Indicator (PI)		Sentiment Quotient (SQ)	
	Rational	**Irrational**	**Rational**	**Irrational**	**Rational**	**Irrational**
			Granger Causality			
Sentiment	0.35	0.68	0.14	0.41	0.21	0.47
Total Return	< 0.01	0.34	< 0.01	0.87	< 0.01	0.76

Note: This table shows the results for the vector autoregression models with exogenous variables (VARX). The models are tested for each sentiment measure (*OI*, *PI*, *SQ*) in the COVID period (02/2020 – 12/2022). The sentiment has been separated into rational and irrational components using an orthogonalization approach based on Ling et al. (2014). The total return of the FTSE Nareit All Equity REITs Index and the sentiment measure each show the lagged results for the past six months. To control for external effects, the control variables *Inflation*, *Unrate*, *Uncertainty*, *Spread,* and *Term* are integrated in the models. Besides the VAR models, the Granger Causality between the independent variable and each sentiment measure is tested, and the respective p-values are presented. ***, **, and * denote significance at the 1%, 5% and 10% level.

Overall, our results in Table 3.6 and 3.7 suggest that pessimistic Twitter sentiment has the highest explanatory power for REIT returns. As hypothesized, irrational (rational) pessimistic Twitter sentiment predicts REIT returns in the pre-COVID (COVID) period. These findings also align with our findings for the higher explanatory power of fundamental variables for REIT-specific Twitter sentiment in the COVID period, compared to the pre-COVID, in Tables 3.2 to 3.5. Based on the 2013 to 2022 period, Paulus et al. (2025) find that pessimistic Twitter sentiment negatively predicts aggregated REIT returns. Our findings suggest that these results are driven by the COVID period and rational Twitter sentiment. The finding of pessimism being a superior sentiment indicator is also in line with previous studies that suggest a negativity bias due to investors being more influenced by negative information (e.g., Rozin & Royzman, 2001).

3.6 Conclusion

Over the last decade, social media such as Twitter has become an important source of information, particularly for individual investors (Eliner & Kobilov, 2023). Paulus et al. (2025) find that the sentiment of Twitter users about REITs can predict aggregated REIT returns. However, the authors fail to provide explanations for why Twitter sentiment can explain REIT returns. One explanation for the explanatory power of Twitter sentiment as a measure of individual investor sentiment is that it reflects fundamentals that impact REITs and their business. Another explanation is that Twitter sentiment reflects noise or irrationality of individual investors that affects REIT returns.

The question of why the sentiment of Twitter users, who are more likely individual investors, can predict returns of the highly specialized and less consumer-oriented REIT industry is interesting, as individual investors tend to focus on large-cap or popular stocks that they are familiar with as consumers (Cookson et al., 2024; Welch, 2022). Furthermore, individual investors tend to invest in attention-grabbing stocks (Barber & Odean, 2008). Historically, REITs were neither of these, and individual investors as a group likely had a lower familiarity with this industry.

However, during the COVID pandemic, positive or negative conditions in commercial real estate markets such as industrial, office, retail, or hotel received more attention in the media and likely resulted in REITs receiving more attention from individual investors. In fact, Ozik et al. (2021) show that stocks, which received a lot of COVID-related media attention, were traded more by individual investors. Coupled with stay-at-home orders, which provided individual investors with more time to research industries and firms, as well as relief cash payments and FinTech innovations that made trading easier, the increased attention of the media on commercial real estate is expected to have led to a higher familiarity of individual investors with REITs in the COVID period. As a result, we hypothesize that in the pre-COVID crisis, the relationship between Twitter sentiment and REIT returns was driven by noise or irrationality. On the other hand, the increased sophistication of individual investors with REITs in the COVID period is expected to lead to fundamentals driving the relationship between Twitter sentiment and REIT returns in this period.

In our empirical investigation, we employ a natural language processing approach to extract optimistic, pessimistic, and overall sentiment measures from REIT-related Twitter comments over the period of 2014 to 2022. Next, we investigate the explanatory power of fundamental variables (macroeconomics, stock, bond, and commercial real estate market) for the Twitter sentiment measures in the pre-COVID and COVID period using an ARDL model. Last, we decompose Twitter sentiment into a rational (fundamentals-based) and irrational component using an orthogonalization approach. We regress the sentiment measures on fundamentals and retain the fitted values (residuals) as the rational (irrational) sentiment component. We then assess the explanatory power of rational and irrational Twitter sentiment for REIT returns in the pre-COVID and COVID period using a VARX model.

In line with expectations, our findings show that the explanatory power of macroeconomic and capital market fundamentals such as inflation, stock market returns, credit risk spread, individual stock investor sentiment, and volatility is higher for optimistic, pessimistic, and overall Twitter sentiment in the COVID period than in the pre-COVID period. We furthermore find that pessimistic Twitter sentiment, as opposed to optimistic or overall sentiment, has the highest explanatory power for REIT returns. In particular, irrational (rational) pessimistic Twitter sentiment predicts REIT returns in the pre-COVID (COVID) period.

Our findings provide evidence for a change in why Twitter sentiment can explain REIT returns. While it mostly captures noise in the pre-COVID period, the improved understanding of REITs by individual investors (Twitter users) led to REIT-specific Twitter sentiment reflecting fundamentals in the COVID period. Our findings suggest that individual investors learned over time and increased their sophistication with regard to REITs. This is in line with previous studies, which

find individual investors also trade on fundamental information (e.g., Kaniel et al., 2012; Kelly & Tetlock, 2013) and learn to improve their abilities (Seru et al., 2010). Future studies may use our findings as a starting point to further investigate the understanding of commercial real estate markets and different property types by individual investors. The sophistication of these types of investors may vary across property types: Individual investors may have a better understanding of property types such as multifamily, hospitality, and retail, while their understanding is more limited with regard to industrial, office, agricultural, or infrastructure. Additionally, while individual investors may have become more familiar with REIT stocks, their understanding of underlying commercial real estate markets, lease types, demand, and supply drivers may still be limited.

Our study furthermore presents Twitter sentiment as a proxy for individual REIT investor sentiment, which has been under-researched in the literature. Future studies may investigate the relationship of institutional and individual REIT investor sentiment in more detail and their individual and joint effects on REIT market volatility, returns, and trading behavior. These studies could investigate, for example, whether individual investors follow the more sophisticated institutional investors or whether these two types of sentiment are driven by different factors.

3.7 References

Ammari, A., Chebbi, K., & Arfa, N. B. (2023). How does the COVID-19 pandemic shape the relationship between Twitter sentiment and stock liquidity of US firms? *International Review of Financial Analysis*, *88*, Article 102633. https://doi.org/10.1016/j.irfa.2023.102633

Ankit & Saleena, N. (2018). An ensemble classification system for Twitter sentiment analysis. *Procedia Computer Science*, *132*, 937–946. https://doi.org/10.1016/j.procs.2018.05.109

Antweiler, W., & Frank, M. Z. (2004). Is all that talk just noise? The information content of internet stock message boards. *The Journal of Finance*, *59*(3), 1259–1294. https://doi.org/10.1111/j.1540-6261.2004.00662.x

Aroul, R., Freybote, J., & Nguyen, A. (2023). Do informed REIT market participants respond to property sector mispricing? *Journal of Property Research*, *40*(4), 311–332. https://doi.org/10.1080/09599916.2023.2222377

Baker, S. R., Bloom, N., & Davis, S. J. (2024). *Economic Policy Uncertainty Index for United States* [USEPUINDXM]. Retrieved October 1, 2025, from https://fred.stlouisfed.org/series/USEPUINDXM

Barber, B. M., Huang, X., Odean, T., & Schwarz, C. (2022). Attention-induced trading and returns: Evidence from Robinhood users. *The Journal of Finance*, *77*(6), 3141–3190. https://doi.org/10.1111/jofi.13183

Barber, B. M., & Odean, T. (2000). Trading is hazardous to your wealth: The common stock investment performance of individual investors. *The Journal of Finance*, *55*(2), 773–806. https://doi.org/10.1111/0022-1082.00226

Barber, B. M., & Odean, T. (2002). Online investors: Do the slow die first? *The Review of Financial Studies*, *15*(2), 455–487. https://doi.org/10.1093/rfs/15.2.455

Barber, B. M., & Odean, T. (2008). All that glitters: The effect of attention and news on the buying behavior of individual and institutional investors. *The Review of Financial Studies*, *21*(2), 785–818. https://doi.org/10.1093/rfs/hhm079

Barber, B. M., Odean, T., & Zhu, N. (2009). Do retail trades move markets? *The Review of Financial Studies*, *22*(1), 151–186. https://doi.org/10.1093/rfs/hhn035

Braun, J., Hausler, J., & Schäfers, W. (2020). Artificial intelligence, news sentiment, and property market liquidity. *Journal of Property Investment & Finance*, *38*(4), 309–325. https://doi.org/10.1108/JPIF-08-2019-0100

Brooks, C., & Tsolacos, S. (2010). *Real estate modelling and forecasting*. Cambridge University Press.

Brown, T. B., Mann, B., Ryder, N., Subbiah, M., Kaplan, J., Dhariwal, P., Neelakantan, A., Shyam, P., Sastry, G., Askell, A., Agarwal, S., Herbert-Voss, A., Krueger, G., Henighan, T., Child, R., Ramesh, A., Ziegler, D. M., Wu, J., Winter, C., . . . Amodei, D. (2020). Language models are few-shot learners. In H. Larochelle, M. Ranzato, R. Hadsell, M. F. Balcan, & H. Lin (Eds.), *Advances in Neural Information Processing Systems* (Vol. 33, pp. 1877–1901). Curran Associates. https://papers.nips.cc/paper_files/paper/2020/file/1457c0d6bfcb4967418bfb8ac142f64a-Paper.pdf

Carstens, R., & Freybote, J. (2019). Tone in REIT financial statements and institutional investments. *Journal of Property Research*, *36*(3), 227–244. https://doi.org/10.1080/09599916.2019.1650802

Chen, H., De, P., Hu, Y., & Hwang, B.-H. (2014). Wisdom of crowds: The value of stock opinions transmitted through social media. *The Review of Financial Studies*, *27*(5), 1367–1403. https://doi.org/10.1093/rfs/hhu001

Clayton, J., Ling, D. C., & Naranjo, A. (2009). Commercial real estate valuation: Fundamentals versus investor sentiment. *The Journal of Real Estate Finance and Economics*, *38*(1), 5–37. https://doi.org/10.1007/s11146-008-9130-6

Cookson, J. A., Lu, R., Mullins, W., & Niessner, M. (2024). The social signal. *Journal of Financial Economics*, *158*, Article 103870. https://doi.org/10.1016/j.jfineco.2024.103870

Cookson, J. A., & Niessner, M. (2020). Why don't we agree? Evidence from a social network of investors. *The Journal of Finance*, *75*(1), 173–228. https://doi.org/10.1111/jofi.12852

Da, Z., Engelberg, J., & Gao, P. (2011). In search of attention. *The Journal of Finance*, *66*(5), 1461–1499. https://doi.org/10.1111/j.1540-6261.2011.01679.x

Das, P. K., Freybote, J., & Marcato, G. (2015). An investigation into sentiment-induced institutional trading behavior and asset pricing in the REIT market. *The Journal of Real Estate Finance and Economics*, *51*(2), 160–189. https://doi.org/10.1007/s11146-014-9490-z

Devlin, J., Chang, M.-W., Lee, K., & Toutanova, K. (2019). BERT: Pre-training of deep bidirectional transformers for language understanding. In J. Burstein, C. Doran, & T. Solorio (Eds.), *Proceedings of the 2019 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies, Volume 1 (Long and Short Papers)* (pp. 4171–4186). Association for Computational Linguistics. https://doi.org/10.18653/v1/N19-1423

Dorle, S., & Pise, N. (2018). Political sentiment analysis through social media. In *2018 Second International Conference on Computing Methodologies and Communication (ICCMC)* (pp. 869–873). IEEE. https://doi.org/10.1109/ICCMC.2018.8487879

Eliner, L., & Kobilov, B. (2023). *To the moon or bust: Do retail investors profit from social media-induced trading?* [Working paper]. https://afajof.org/management/viewp.php?n=64120

Freybote, J., & Seagraves, P. A. (2017). Heterogeneous investor sentiment and institutional real estate investments. *Real Estate Economics*, *45*(1), 154–176. https://doi.org/10.1111/1540-6229.12132

Granger, C. W. J. (1969). Investigating causal relations by econometric models and cross-spectral methods. *Econometrica*, *37*(3), 424–438. https://doi.org/10.2307/1912791

Gu, C., & Kurov, A. (2020). Informational role of social media: Evidence from Twitter sentiment. *Journal of Banking & Finance*, *121*, Article 105969. https://doi.org/10.1016/j.jbankfin.2020.105969

Han, B., Hirshleifer, D., & Walden, J. (2022). Social transmission bias and investor behavior. *Journal of Financial and Quantitative Analysis*, *57*(1), 390–412. https://doi.org/10.1017/S0022109021000077

Hausler, J., Ruscheinsky, J., & Lang, M. (2018). News-based sentiment analysis in real estate: A machine learning approach. *Journal of Property Research*, *35*(4), 344–371. https://doi.org/10.1080/09599916.2018.1551923

Hirshleifer, D. (2020). Presidential address: Social transmission bias in economics and finance. *The Journal of Finance*, *75*(4), 1779–1832. https://doi.org/10.1111/jofi.12906

Hoesli, M., Lizieri, C., & MacGregor, B. (2008). The inflation hedging characteristics of US and UK investments: A multi-factor error correction approach. *The Journal of Real Estate Finance and Economics*, *36*(2), 183–206. https://doi.org/10.1007/s11146-007-9062-6

Jalil, A., & Rao, N. H. (2019). Time series analysis (stationarity, cointegration, and causality). In B. Özcan & I. Öztürk (Eds.), *Environmental Kuznets Curve (EKC)* (pp. 85–99). Academic Press. https://doi.org/10.1016/B978-0-12-816797-7.00008-4

Jame, R., Johnston, R., Markov, S., & Wolfe, M. C. (2016). The value of crowdsourced earnings forecasts. *Journal of Accounting Research*, *54*(4), 1077–1110. https://www.jstor.org/stable/24738215

Jones, R. (2022). *Internet & text slang dictionary*. Retrieved August 22, 2022, from https://www.noslang.com/dictionary/

Kaniel, R., Liu, S., Saar, G., & Titman, S. (2012). Individual investor trading and return patterns around earnings announcements. *The Journal of Finance*, *67*(2), 639–680. https://doi.org/10.1111/j.1540-6261.2012.01727.x

Karampatsas, N., Malekpour, S., Mason, A., & Mavis, C. P. (2023). Twitter investor sentiment and corporate earnings announcements. *European Financial Management*, *29*(3), 953–986. https://doi.org/10.1111/eufm.12384

Karanikola, A., Davrazos, G., Liapis, C. M., & Kotsiantis, S. (2023). Financial sentiment analysis: Classic methods vs. deep learning models. *Intelligent Decision Technologies*, *17*(4), 893–915. https://doi.org/10.3233/IDT-230478

Kelly, E. K., & Tetlock, P. C. (2013). How wise are crowds? Insights from retail orders and stock returns. *The Journal of Finance*, *68*(3), 1229–1265. https://doi.org/10.1111/jofi.12028

Kumar, A. (2009). Who gambles in the stock market? *The Journal of Finance*, *64*(4), 1889–1933. https://doi.org/10.1111/j.1540-6261.2009.01483.x

Lin, C. Y., Rahman, H., & Yung, K. (2009). Investor sentiment and REIT returns. *The Journal of Real Estate Finance and Economics*, *39*(4), 450–471. https://doi.org/10.1007/s11146-008-9115-5

Ling, D. C., Naranjo, A., & Scheick, B. (2014). Investor sentiment, limits to arbitrage and private market returns. *Real Estate Economics*, *42*(3), 531–577. https://doi.org/10.1111/1540-6229.12037

Liu, Y., Ott, M., Goyal, N., Du, J., Joshi, M., Chen, D., Levy, O., Lewis, M., Zettlemoyer, L., & Stoyanov, V. (2019). *RoBERTa: A robustly optimized BERT pretraining approach*. arXiv. https://doi.org/10.48550/arxiv.1907.11692

Manning, C. D., Raghavan, P., & Schütze, H. (2008). The term vocabulary and postings lists. In *Introduction to information retrieval* (pp. 18–44). Cambridge University Press. https://doi.org/10.1017/CBO9780511809071

Newey, W. K., & West, K. D. (1987). A simple, positive semi-definite, heteroskedasticity and autocorrelation consistent covariance matrix. *Econometrica*, *55*(3), 703–708. https://doi.org/10.2307/1913610

Ozik, G., Sadka, R., & Shen, S. (2021). Flattening the illiquidity curve: Retail trading during the COVID-19 lockdown. *Journal of Financial and Quantitative Analysis*, *56*(7), 2356–2388. https://doi.org/10.1017/S0022109021000387

Paulus, N. M., Lautenschlaeger, L., & Schaefers, W. (2025). Social media and real estate: Do Twitter users predict REIT performance? *Journal of Real Estate Research*, *47*(3), 322–355. https://doi.org/10.1080/08965803.2024.2340915

Ranco, G., Aleksovski, D., Caldarelli, G., Grcar, M., & Mozetic, I. (2015). The effects of Twitter sentiment on stock price returns. *PLoS One*, *10*(9), Article e0138441. https://doi.org/10.1371/journal.pone.0138441

Reboredo, J. C., & Ugolini, A. (2018). The impact of Twitter sentiment on renewable energy stocks. *Energy Economics*, *76*, 153–169. https://doi.org/10.1016/j.eneco.2018.10.014

Rozin, P., & Royzman, E. B. (2001). Negativity bias, negativity dominance, and contagion. *Personality and Social Psychology Review*, *5*(4), 296–320. https://doi.org/10.1207/S15327957PSPR0504_2

Ruscheinsky, J. R., Lang, M., & Schäfers, W. (2018). Real estate media sentiment through textual analysis. *Journal of Property Investment & Finance*, *36*(5), 410–428. https://doi.org/10.1108/JPIF-07-2017-0050

Schmidt, D. (2020). Stock market rumors and credibility. *The Review of Financial Studies*, *33*(8), 3804–3853. https://doi.org/10.1093/rfs/hhz120

Seru, A., Shumway, T., & Stoffman, N. (2010). Learning by trading. *The Review of Financial Studies*, *23*(2), 705–739. https://doi.org/10.1093/rfs/hhp060

Shiller, R. J. (2017). Narrative economics. *American Economic Review*, *107*(4), 967–1004. https://doi.org/10.1257/aer.107.4.967

Sprenger, T. O., Sandner, P. G., Tumasjan, A., & Welpe, I. M. (2014). News or noise? Using Twitter to identify and understand company-specific news flow. *Journal of Business Finance & Accounting*, *41*(7 & 8), 791–830. https://doi.org/10.1111/jbfa.12086

Sul, H. K., Dennis, A. R., & Yuan, L. (2017). Trading on Twitter: Using social media sentiment to predict stock returns. *Decision Sciences*, *48*(3), 454–488. https://doi.org/10.1111/deci.12229

Tan, S. D., & Tas, O. (2021). Social media sentiment in international stock returns and trading activity. *Journal of Behavioral Finance*, *22*(2), 221–234. https://doi.org/10.1080/15427560.2020.1772261

Tang, V. W. (2018). Wisdom of crowds: Cross-sectional variation in the informativeness of third-party-generated product information on Twitter. *Journal of Accounting Research*, *56*(3), 989–1034. https://doi.org/10.1111/1475-679X.12183

Welch, I. (2022). The wisdom of the Robinhood crowd. *The Journal of Finance*, *77*(3), 1489–1527. https://doi.org/10.1111/jofi.13128

Zeitun, R., Rehman, M. U., Ahmad, N., & Vo, X. V. (2023). The impact of Twitter-based sentiment on US sectoral returns. *North American Journal of Economics & Finance*, *64*, Article 101847. https://doi.org/10.1016/j.najef.2022.101847

4 Signals or Stories? ESG Sentiment and Investor Reactions in REIT Corporate Disclosure

4.1 Abstract

Purpose: This paper examines whether ESG-related sentiment in corporate disclosures of U.S. equity REITs is reflected in stock returns. It focuses on how investors respond to ESG language across disclosure channels, ESG pillars, and REIT subsectors.

Design/Methodology/Approach: We apply fine-tuned RoBERTa models to identify ESG-related text and classify sentiment in 10-Ks, 10-Qs, and standalone ESG reports for 174 U.S. equity REITs from 2016–2025. Event-study regressions relate pillar-specific Net Positive Tone measures to cumulative abnormal returns over multiple event windows.

Findings: Aggregate ESG sentiment shows limited explanatory power for abnormal returns, whereas disaggregated analyses reveal systematic differences across pillars and sectors. Environmental sentiment, especially in operationally intensive property types, exhibits the most robust positive association with cumulative abnormal returns, while social and governance sentiment display weaker or context-dependent effects.

Research Limitations/Implications: Results are conditional on the U.S. equity REIT universe, the 2016–2025 sample period, and the chosen LLM-based classification approach. Future research could examine alternative sentiment models, cross-country REIT samples, and interactions with regulatory changes in ESG disclosure.

Practical Implications: For investors, the findings highlight that ESG language in REIT disclosures is most informative when focusing on environmental and, to a lesser extent, social sentiment in economically exposed sectors. For managers, they stress the importance of credible, financially linked ESG communication rather than generic ESG messaging.

Originality/Value: The study is the first to combine LLM-based ESG sentiment extraction with event-study regressions for REIT disclosures, and to compare mandatory and voluntary reports at the pillar and sub-sector level. It shows that ESG communication is processed selectively rather than as a homogeneous signal in real estate capital markets.

Keywords: Real Estate Investment Trusts, Corporate Disclosures, Sentiment, ESG, Large Language Models, ESG reports

4.2 Introduction

In recent decades, environmental, social, and governance (ESG) considerations have become an increasingly important driver of corporate strategy and investment decisions worldwide. This development is particularly relevant in the real estate and built environment sector, which accounts for a substantial share of global energy-related greenhouse gas emissions. With a market capitalization of around 1.47 trillion USD (Nareit, 2025a) and being closely connected to direct real estate markets, Real Estate Investment Trusts (REITs) play a central role at the intersection of financial markets and the sustainability challenges the real estate sector must face (Morri et al., 2024; Robinson & McIntosh, 2022), including the decarbonization of the economy (OECD, 2025; Tiwari & Miao, 2022). Their portfolio and investment decisions directly influence the environmental and social performance of the built environment, while their securities are continuously valued on public markets (Fan et al., 2024).

From an environmental perspective, REITs are directly exposed to both physical and climate transition risks. Physical risks such as extreme weather events, rising sea levels, and temperature fluctuations can affect asset values, disrupt operations, and reduce net operating income (NOI). At the same time, transition risks arising from stricter environmental regulations, carbon pricing, and changing tenant and investor preferences can increase compliance costs and capital requirements and potentially impact income streams and valuations (Contat et al., 2024). The social dimension of ESG has become increasingly important as well (Robinson & McIntosh, 2022; Zheng & Lee, 2026), particularly in connection with the heightened focus on diversity, equity, and inclusion (DEI) in the U.S. In the REIT context, social considerations operate on multiple levels. On the tenant side, factors such as health and safety, affordability, accessibility, and tenant satisfaction influence occupancy rates, lease renewals, and reputation. At the company level, workforce diversity, work practices, and employee well-being affect productivity, employee retention, and corporate culture. However, social factors are still difficult to measure and compare across firms. Therefore, comprehensive disclosure plays a crucial role in reporting on S. Governance is the third pillar of ESG and is particularly important for REITs due to their regulatory structure, agency issues, and dependence on debt capital. Effective governance mechanisms include transparent reporting, robust risk management, and conflict resolution frameworks. By doing so, companies can mitigate agency issues, reduce legal and regulatory risks, and protect themselves from reputational damage. Weak governance, on the other hand, can increase operational risks and undermine investor confidence (Zheng & Lee, 2026).

Given the growing importance of ESG information for investors, regulatory authorities have increasingly focused on disclosure requirements. The U.S. Securities and Exchange Commission (SEC) has recognized the importance of ESG-related risks and initially introduced mandatory ESG

disclosure in 2024 (SEC, 2024). Although this decision has since been challenged and partially reversed due to shifted priorities at the beginning of 2025, it remains a hot topic in the regulatory debate (Tomlinson, 2025). In parallel, industry organizations such as the National Association of Real Estate Investment Trusts (Nareit) have emphasized the importance of ESG transparency by providing reporting guidelines and frameworks tailored to REITs (Nareit, 2025b). They argue that ESG best practices are not only beneficial to the environment and society but also improve long-term business performance and capital-market access.

Despite this, there is still limited empirical evidence on how ESG-related information in textual form is processed and evaluated by investors in the REIT sector. Existing studies are partly based on aggregated ESG scores, ratings, or quantitative performance indicators (e.g., Aroul et al., 2022; Morri et al., 2024; Zheng & Lee, 2026), which may obscure important differences between environmental, social, and governance issues. Other REIT research shows that the individual pillars matter and behave differently: environmental issues are closely linked to asset characteristics and regulatory and transition risks (Fan et al., 2024; Morri et al., 2024), social issues relate to interactions with tenants, employees, and communities and affect operating results and reputation (Carnini Pulino et al., 2022; Kim & Li, 2021), and corporate governance is strongly shaped by the REIT-specific institutional framework and regulatory structure (Kohl & Schaefers, 2012; Zheng & Lee, 2026). Likewise, prior work on REIT disclosure documents that the narrative tone in financial reports contains value-relevant information for investors. The tone of earnings calls, MD&A sections, and forward-looking passages predicts short-term announcement returns and future operating performance (Carstens & Freybote, 2021; Doran et al., 2012; Koelbl, 2020). However, the specific role of ESG-related language and sentiment within these disclosures has been largely unexplored. It remains unclear whether investors distinguish between ESG statements in regulated financial reports and those in voluntary ESG or sustainability reports, whether aggregate ESG sentiment suffices, and how market reactions vary by ESG pillar and REIT subsector.

This study addresses these gaps by examining the value relevance of ESG-related sentiment in REIT disclosures. Focusing on U.S. equity REITs, we analyze ESG sentiment derived from both mandatory Form 10-K and 10-Q filings and voluntary ESG reports, allowing us to compare investor reactions across different types of disclosures. Specifically, we look at the cumulative abnormal returns (CARs) on the respective disclosure dates. We take a disaggregated perspective on asset-class-level and examine the environmental, social, and governance pillars separately to capture potential heterogeneities in investor reactions. To filter for ESG-specific sections and to extract their sentiment, we use an advanced large language model (LLM), which allows for a more nuanced language interpretation than traditional lexicon-based methods.

We aim to make several contributions to real estate literature with this study. First, it expands the growing body of research on ESG and REIT performance by shifting the focus from ESG performance scores and ratings to ESG communication and sentiment in textual disclosures. Second, it contributes to the literature on corporate disclosure and capital market reactions by providing evidence on how investors process ESG narratives in both mandatory and voluntary reporting. Third, by analyzing the pillar and subsector level, the study highlights the heterogeneity of ESG across different REIT asset classes and shows that environmental, social, and governance information are not uniformly priced. To our knowledge, we are the first to analyze ESG in corporate disclosure on a subsector level in the context of REITs. The rest of the paper is structured as follows: Section 4.3 provides an overview of the relevant literature and the development of the hypotheses. Section 4.4 describes the data and methodology. Section 4.5 presents the empirical results, and Section 4.6 concludes the paper.

4.3 Literature Review

4.3.1 Overall Effect of ESG in Corporate Disclosure

ESG information has evolved into an essential component of corporate disclosure, as it is now systematically used by investors, analysts, and other stakeholders to assess firm value, risk, and long-term performance. It can reduce a company's cost of capital, attract ESG-oriented investors, and thereby strengthen the link between ESG ratings and reports and financial performance. Additionally, it is linked to growth and investment opportunities (Chen & Xie, 2022; Chiang et al., 2019; Dhaliwal et al., 2012). Strong ESG or environmental performance is also associated with more flexible funding conditions, higher credit ratings, and a positive relation to firm value (Eichholtz et al., 2019; Feng & Wu, 2023). ESG report characteristics such as readability, tone, length, and the use of numerical or forward-looking content help explain why some reports improve the information environment while others do not. More readable and higher-quality documents are associated with higher firm value and more accurate analyst forecasts (Huang et al., 2025; Muslu et al., 2019).

From a market-based perspective, ESG information is reflected in share prices and risk measures. Positive ESG news reduces stock-price crash risk, particularly for firms with low analyst coverage and weak transparency, and tends to generate moderate positive abnormal returns, while negative ESG news triggers much stronger negative reactions, with these effects shaped by firms' historical ESG records (Dorfleitner & Zhang, 2024; Jeong et al., 2025; Yu et al., 2023). Higher ESG ratings and active reporting about physical climate risks have also been associated with a lower risk of stock-price crashes, suggesting that ESG may offer partial protection against extreme negative events (Feng et al., 2022; Hunjra et al., 2020; Schiemann & Sakhel, 2019).

However, several recent studies caution against treating ESG as a solely positive driver. The disclosures can also have negative effects when claims are not backed by genuine performance improvements. Unsubstantiated or strategic greenwashing may lower firm value and trigger investor skepticism, especially in markets with heightened ESG scrutiny (Cerciello et al., 2023; Chacon et al., 2025; Fatemi et al., 2018). Very high ESG commitments can coincide with lower firm value, weaker operating cash flows, and heightened risk, suggesting overinvestment or immaterial ESG spending that is not compensated for by cash-flow gains. Beyond an optimal ESG level, further improvements in numerical ESG scores can be associated with weaker return performance (Kakinuma, 2025; Morri et al., 2024). Additionally, in the equity market, ESG-rated REITs have been found to underperform non-rated peers in returns (Neo et al., 2025).

Overall, the impact of ESG activities and their disclosure on firm value and risk is far from uniform and remains mixed and context-dependent across firms and information environments (Gillan et al., 2021). Meta-analytic evidence confirms that, on average, ESG engagement is associated with improved financial performance in real estate, albeit with substantial variation across market segments and pillars (Aroul et al., 2022; Morri et al., 2024; Zheng & Lee, 2026). Higher ESG quality also appears to work through real-economy channels, as better ESG performance is associated with higher rents and ESG-driven tenant demand for green buildings (Liu et al., 2025).

4.3.2 Sentiment Analysis as a Tool to Assess ESG Performance

One way to assess ESG performance in firm communication is to quantify qualitative information using sentiment analysis. Prior work shows that textual sentiment from news and social media predicts REIT market outcomes and that managerial tone in earnings calls, MD&A, and financial statements carries information about returns and operating performance (Braun et al., 2020; Carstens & Freybote, 2021; Doran et al., 2012; Hausler et al., 2018; Koelbl, 2020; Paulus et al., 2025; Ruscheinsky et al., 2018). Despite this, the specific role of ESG-related sentiment within REIT disclosures remains largely unexplored.

For non-REITs, however, ESG-specific sentiment is clearly priced. Positive ESG news sentiment lowers stock price crash risk, especially when traditional disclosure and analyst coverage are weak, and ESG news sentiment generates abnormal returns, with negative ESG news having a stronger effect than positive news (Dorfleitner & Zhang, 2024; Jeong et al., 2025; Yu et al., 2023). Likewise, the tone of ESG-related language in 10-K filings significantly affects stock returns around disclosure dates, indicating that investors do not treat ESG language and ESG-related news as background noise but actively trade on these sentiment signals (Ignatov, 2023). Therefore, we address this gap in the REIT context by analyzing the effect of ESG-related sentiment in REIT 10-K and 10-Q filings as well as standalone sustainability reports, as supported and recommended by the Nareit (Nareit, 2025b). Our first hypothesis is:

H1: *ESG-related sentiment in REIT corporate disclosure is associated with CARs of U.S. REITs.*

4.3.3 A Disaggregated View of the ESG Impact

The overall assessment of ESG has led to divergent results in past studies, partly because ESG combines dimensions with differing financial relevance. Recent work therefore emphasizes a pillar- or component-level perspective to understand ESG impacts (Erol et al., 2023; Fan et al., 2024; Han et al., 2025; Zheng & Lee, 2026). For the environmental pillar, evidence is mixed. Environmental initiatives such as certifications and emissions reductions can support lower financing costs and, in the REIT context, "greenness" is associated with stronger operating performance, but environmental efforts can also be costly and are not guaranteed to translate into higher profitability or returns (Bauer & Hann, 2010; Brounen & Marcato, 2018; Fan et al., 2024; Morri et al., 2021, 2024). In broader corporate-finance samples, higher environmental- and social-risk scores may reduce corporates' financial stability and raise their default risks, incurring default costs (Cohen, 2023).

Besides the E, social performance emerges as the most consistently positive driver of financial outcomes in the context of ESG. Firms disclosing strong social metrics partly outperform peers, earn valuation premia and improved investor sentiment, and long-horizon institutions systematically underweight stocks with negative E and S flags (Brounen & Marcato, 2018; Carnini Pulino et al., 2022; Nofsinger et al., 2019). In REITs, social performance is positively associated with future returns, firm performance, and has a robust influence on performance in meta-analytic evidence, with E-investing and especially S-investing practices identified as the main financial performance drivers (Erol et al., 2023; Fan et al., 2024; Zheng & Lee, 2026). The governance pillar generally shows the most nuanced and context-dependent patterns. In non-REIT samples, there is a positive relationship between governance and corporate profitability, however only observed in firms with weak governance (Kim & Li, 2021). In early REIT studies, governance was found to matter to capital market participants (Kohl & Schaefers, 2012), while now governance scores are often insignificant as stand-alone performance predictors and can even be associated with adverse outcomes (Brounen & Marcato, 2018; Erol et al., 2023; Morri et al., 2024; Zheng & Lee, 2026).

The evidence in sentiment analysis shows that overall ESG news sentiment is positively related to corporate performance (driven by social and governance news sentiment), whereas environmental news sentiment is insignificant (Han et al., 2025). These findings support the notion that both the strength and direction of ESG relationships are shaped by disaggregating ESG into its components. In both performance and sentiment studies, social and environmental metrics tend to show the clearest economic effects, whereas governance often behaves more as a conditioning or moderating factor than a primary return driver, especially in REITs (Erol et al.,

2023; Fan et al., 2024; Morri et al., 2021, 2024; Zheng & Lee, 2026). This motivates our second hypothesis:

H2: *Sentiment regarding individual ESG pillars holds more explanatory power toward REIT returns than aggregate ESG sentiment.*

4.3.4 Heterogeneity of ESG Sentiment in REIT Subsectors

Different real estate sectors serve different user groups and have distinct operational and technological profiles, so aggregating them can easily obscure heterogeneous ESG effects. Prior work shows that the information content of disclosure tone is property-type specific. Forward-looking tone in REIT financial statements predicts subsequent returns for residential and retail REITs, behaves differently for office REITs, and carries no informative value for industrial REITs, indicating that investors interpret similar linguistic signals differently across asset classes (Carstens & Freybote, 2021). Liang et al. (2021) likewise document that also ESG performance varies systematically with REIT portfolio focus. REITs that concentrate on less transparent, service-intensive assets such as hotels and hospitals tend to exhibit weaker environmental and social responsibility scores and lower overall ESG performance, while commercial offices and parts of the residential and industrial stock are more transparent on environmental performance because they are more often subject to professional environmental assessment schemes and regulatory frameworks. Governance scores show no systematic differences across asset types, consistent with governance being largely determined by the common REIT regulatory and corporate-law framework rather than by the specific underlying assets (Liang et al., 2021). Additionally, Jeong et al. (2025) find that ESG news sentiment has the strongest positive association with performance in industries with pronounced ESG concerns and that social and governance news sentiment dominate these effects, suggesting that investors selectively amplify the ESG dimensions that are most material for a given sector.

Taken together, these findings imply that the three ESG pillars are unlikely to be uniformly priced across REIT subsectors. From an investor viewpoint, environmental sentiment should be most relevant where carbon and energy risks dominate and signal future retrofit and capex needs (e.g., residential, office, industrial, lodging, health care, data centers), social sentiment where human-service intensity and direct stakeholder contact are highest (e.g., lodging, health care, retail), and governance mainly as a background condition under a common REIT regime that becomes critical primarily in cases of visible governance weaknesses. This leads to our third and last hypothesis:

H3: *ESG-related sentiment in REIT corporate disclosure has varying impacts across REIT subsectors.*

4.4 Data and Methodology

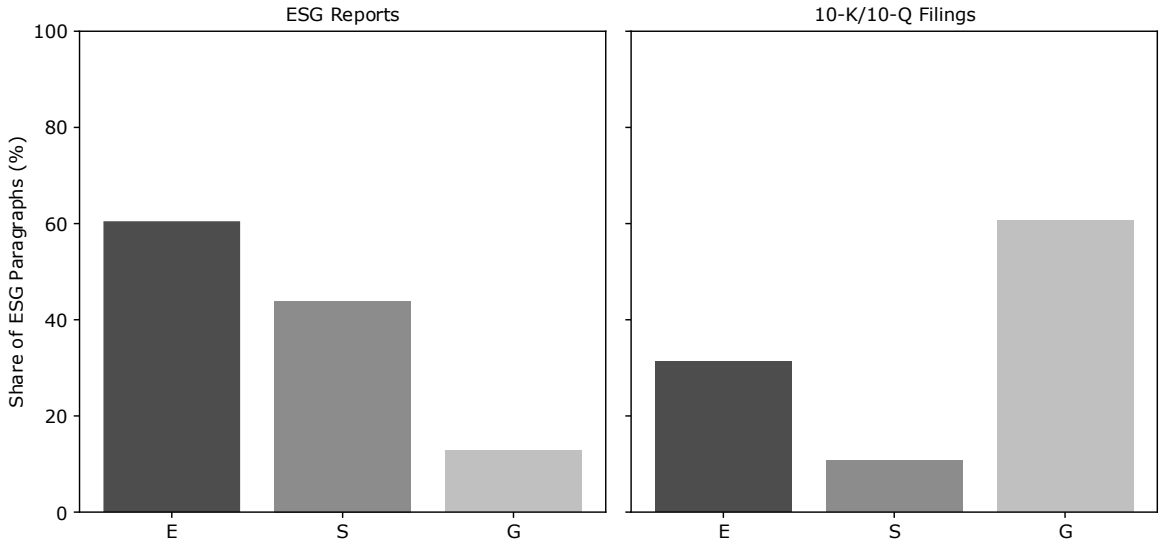
The primary data we investigate contains the textual data from yearly and quarterly financial statements mandated by the SEC (10-K, 10-Q filings) for publicly traded U.S. equity REITs and hand-collected standalone ESG, sustainability, and Corporate Social Responsibility (CSR) reports (summarized as ESG reports), as well as financial data for each individual REIT. Additionally, we utilize macroeconomic and market-specific metrics to control for the current economic environment. For our analysis, we consider the data of 174 REITs that were Nareit index constituents at any time between 2016 and 2025. This period is selected due to the limited availability of ESG reports before 2016, where the relevance of ESG-related reporting was comparably low to today. In total, we have gathered ~ 1,500 Form 10-K, ~ 3,600 Form 10-Q, and ~ 400 ESG reports.

4.4.1 Filtering for ESG Relevance

To classify the derived text into the three pillars E, S, and G, we utilize a task-specifically fine-tuned Robustly Optimized BERT Pretraining Approach (RoBERTa) model. RoBERTa was originally developed by Liu et al. (2019) and is based on the Bidirectional Encoder Representations from Transformers (BERT) model introduced by Devlin et al. (2019). By using extensively pre-trained models, we are minimizing the need for a training dataset, as the learning process of the overall textual understanding has already happened. Therefore, it is only necessary to fine-tune the models for the desired task. For fine-tuning, we utilize the environmental_2k, social_2k, and governance_2k datasets made available on Hugging Face by Schimanski et al. (2024). These datasets each consist of approximately 2,000 sentences regarding their specific ESG pillar. The resulting ESG-RoBERTa is then used for a multiclass classification task where each paragraph is labeled regarding its affiliation with the three individual parts of ESG. Also, a multiple or no classification (labeled as neutral) of the paragraphs is possible.

Figure 4.1 below shows the number of paragraphs classified as E, S, and G for both the ESG reports and 10-K filings. Across both disclosure types, E-related sections account for a larger share of ESG-specific text than S-related sections. In ESG reports, G-related segments make up the smallest share, whereas in mandatory SEC filings, they represent the largest share. This pattern is plausible since standalone ESG reports are an additional, largely voluntary channel in which firms mainly highlight environmental and social initiatives (Nareit, 2025b), while 10-K and 10-Q filings are tightly regulated documents whose core content (board structure, risk factors, controls, legal matters, financial policies) is inherently governance-oriented, resulting in a much higher proportion of G-related language.

Figure 4.1 – Share of ESG Paragraphs in the ESG Reports and Form 10-Ks/10-Qs



4.4.2 Derivation of Tone Measures

For the task of sentiment classification, we again utilize the RoBERTa model. This time, we fine-tune the model using the Financial PhraseBank dataset by Malo et al. (2014). This dataset consists of ~ 4,800 messages from a financial context that are each prelabeled as positive, negative, or neutral sentiment. With the resulting fine-tuned Sentiment-RoBERTa model, we classify each paragraph into one of the three sentiment classes.

After classifying the tone of each paragraph in the Form 10-K and 10-Q filings as well as the ESG reports, we aggregate these labels to construct the Net Positive Tone (*NPT*) for each document. The *NPT* reflects the relative predominance of positive paragraphs by comparing the difference between the numbers of positive and negative paragraphs to the total number of paragraphs, as shown in Equation 4.1:

$$NPT_t = \frac{\sum_{i=1}^{I} \text{positive paragraph}_{i,t} - \sum_{j=1}^{J} \text{negative paragraph}_{j,t}}{\text{total number of paragraphs}_t} \tag{4.1}$$

where i and j index positive and negative paragraphs, and t denotes the publication date of the report. *NPT* ranges from -1 to 1, where values below 0 indicate a tendency toward a generally negative tone, and values above 0 indicate a tendency toward a more positive overall tone.

4.4.3 Event Study and Regression Analysis

For our regression analysis, we employ an event study methodology based on an ordinary least squares (OLS) regression framework. The dependent variable is the CAR for each firm around the publication dates of the reports. For the 10-K and 10-Q filings, we use the EDGAR timestamps specified in the disclosures, while for ESG reports, we use the publication dates indicated in press releases on the companies' websites. The CARs are calculated using the historical mean-adjusted

return model, which assumes that a firm's normal return can be approximated by its average return over an estimation window. We define the estimation window as the period from [-120, -30] trading days before the event to avoid any influence from the event itself. We consider four event windows [-1,1], [-3,3], [-5,5], and [-1,10], capturing short-, medium-, and longer-term market reactions around the event (Koelbl, 2020). To ensure robust statistical inference, we use two-way clustered standard errors to account for both serial correlation within firms and cross-sectional correlation across firms at specific points in time. This adjustment mitigates potential biases arising from contemporaneous market shocks. The abnormal return for firm i on day t is given by:

$$AR_{i,t} = R_{i,t} - \bar{R}_i \tag{4.2}$$

where $R_{i,t}$ denotes the actual return of firm i on day t, and $\bar{R}_i$ is the average return of the same firm over the estimation window. The CARs are then calculated by summing the abnormal returns over the event window:

$$CAR_{i,t}(\tau_1, \tau_2) = \sum_{t=\tau_1}^{\tau_2} AR_{i,t+\tau} \tag{4.3}$$

Here, t denotes the publication date, while τ denotes the trading day relative to the publication date. To analyze the relationship between sentiment and market reactions, we regress the CARs on the *NPT* of the reports and include a set of control variables. The regression specification is:

$$CAR_{i,t}(\tau_1, \tau_2) = \beta_1 NPT_{i,t} + \beta_2 NPT_{i,t-1} + \gamma' X_{i,t} + \alpha_i + \delta_t + \varepsilon_{i,t} \tag{4.4}$$

where $CAR_{i,t}$ is the cumulative abnormal return for firm i around the report publication at time t, $NPT_{i,t}$ denotes the tone of the disclosure, $NPT_{i,t-1}$ denotes its one-year lag, $X_{i,t}$ represents the set of control variables, α_i and δ_t denote firm and time fixed effects, and $\epsilon_{i,t}$ is the error term.

We use a range of macroeconomic, market-based, and firm-level control variables. Macroeconomic conditions are captured by the quarterly U.S. gross domestic product (*GDP*) growth, measured as the lagged change in log *GDP*, and by the monthly unemployment rate (*Unrate*), included as a rolling standardized z-score. Market conditions are reflected by the lagged daily change in the S&P 500 (*SP500*) log return, the lagged daily change in the effective federal funds rate (*DFF*), the lagged monthly change in the Moody's Seasoned Baa–Aaa corporate bond yield spread (*Spread*), the lagged standardized z-score of the news-based Economic Policy Uncertainty Index (*Uncertainty*) (Baker et al., 2016), and the lagged standardized GDELT z-score (*GDELT*) (Leetaru & Schrodt, 2025). Firm-level characteristics are controlled for using lagged quarterly leverage (*Leverage*), lagged log market capitalization (*MCAP*), and the lagged log market-to-book ratio (*MTB*). All variables are transformed and standardized as reported in Table 4.1, which also provides their descriptive statistics.

Table 4.1 – Descriptive Statistics

Variable	Transformation	Mean	Median	Std	Min	Max
NPT	none	-0.015	-0.010	0.033	-1.000	0.083
NPT_{t-1}	none	-0.014	-0.011	0.021	-0.270	0.037
SP500	Δln(*SP500*)	0.000	0.000	0.013	-0.165	0.130
DFF	Δ(*DFF*)	0.001	0.000	0.040	-0.850	0.750
GDELT	z-score	-0.029	-0.005	1.062	-8.248	6.797
GDP	Δln(*GDP*)	0.004	0.000	0.014	-0.086	0.084
Unrate	z-score	0.122	-0.772	4.867	-3.065	48.719
Spread	Δ(*Spread*)	-0.007	-0.010	0.085	-0.250	0.440
Uncertainty	z-score of ln(1+*Uncertainty*)	0.215	0.184	1.488	-3.275	7.550
Leverage	none	0.503	0.480	0.176	0.000	1.482
MCAP	ln(*MCAP*)	21.895	21.934	1.969	14.169	35.318
MTB	ln(*MTB*)	0.629	0.581	1.392	-4.921	16.736

Note: The table presents descriptive statistics for the sample period 2016–2025. *NPT* is the sentiment measure; NPT_{t-1} is the lagged sentiment measure. *SP500* is the lagged S&P 500 return, DFF is the change in the effective federal funds rate, and *GDELT* is the GDELT-based news measure. *GDP* denotes the (lagged) growth rate of U.S. gross domestic product, *Unrate* is the U.S. unemployment rate, and *Spread* is the Moody's Seasoned Baa–Aaa corporate bond yield spread. *Uncertainty* is the news-based economic policy uncertainty measure. *Leverage* is the firm leverage control, *MCAP* is market capitalization, and *MTB* is market-to-book. Detailed variable definitions and transformations are provided in Section 4.4.

Overall, daily S&P 500 returns and changes in the federal funds rate are close to zero with moderate volatility, consistent with typical short-term market dynamics. The standardized *GDELT*, *Unrate*, and *Uncertainty* measures are close to zero on average, but exhibit pronounced tails, indicating periods of heightened news sentiment, labor-market stress, and macro uncertainty. *GDP* growth and *Spread* vary less, reflecting smoother macroeconomic movements relative to financial conditions. At the firm level, REITs display moderately similar leverage, some dispersion in size (*MCAP*), and substantial heterogeneity in valuations (*MTB*). Finally, we also consider time fixed effects to absorb major macroeconomic shock phases such as the COVID-19 pandemic, the start of the invasion of Ukraine, and the interest rate hike cycle, thereby capturing general changes in investor behavior that are not related to company-level disclosures.

4.5 Results

Table 4.2 shows the results for the 10-Ks, 10-Qs, and ESG reports on an overall and joint ESG level. The findings suggest that the publication of official disclosure mandatory by the SEC is associated with positive and significant mean CARs in the majority of the event windows ([-1,1]; [-3,3]; [-5,5]). This suggests that investors acknowledge the informational value of corporate disclosure and the managerial sentiment within those reports. This pattern is consistent with the established literature documenting that annual reports reduce information asymmetry and uncertainty and contain investor-relevant qualitative signals. (Carstens & Freybote, 2021; Doran et al., 2012; Koelbl, 2020). The sentiment in voluntary ESG reports, on the other hand, is not significant. Also, neither the aggregate ESG sentiment in 10-K/10-Q filings nor the sentiment in voluntary reports shows a clear overall association with the CARs. A potential interpretation is that investors treat the publication of core financial filings itself and the underlying managerial sentiment as a value-relevant event, while high-level ESG language, especially when reported

outside the regulated financial statements, appears too noisy at the aggregate level to shift prices. Another explanation could be that separate ESG reporting might feel like greenwashing or a sign of overinvesting in ESG matters by the companies to the investors. We therefore reject our H1, which stated that ESG-related sentiment in REIT corporate disclosure is associated with the CARs. This rejection again motivates the subsequent disaggregated analysis by ESG pillar and subsector.

Table 4.2 – Regression Results: Overall and Overall ESG Sentiment

	10-K/10-Q vs. ESG reports							
	[-1,1]	**[-3,3]**	**[-5,5]**	**[-1,10]**	**[-1,1]**	**[-3,3]**	**[-5,5]**	**[-1,10]**
	Overall - 10-K/10-Q				**Overall - ESG Reports**			
NPT	0.0733*	0.1192**	0.1679**	0.1424	0.0036	-0.1169	-0.0705	-0.0295
NPT_{t-1}	0.0886	0.0630	0.1581	0.0173	-0.0194	0.0069	-0.2015	-0.0649
SP500	0.4582***	0.9327***	0.8166***	0.9531***	0.9175**	1.0789*	1.0219	1.2462**
DFF	0.0084	-0.0053	0.0600	0.0725*	-0.0325	-0.0219	-0.1163**	-0.0148
GDELT	0.0013	0.0022	0.0004	-0.0033	-0.0005	0.0019	-0.0013	-0.0031
GDP	0.4125***	0.6651***	0.7626***	0.3490*	1.1218*	2.3987***	2.1013**	2.6361***
Unrate	0.0002	-0.0001	-0.0002	-0.0020**	0.0006	-0.0005	-0.0024	-0.0020
Spread	0.0448*	0.0355	0.0976*	0.2088***	0.0763	0.2088**	0.2160	0.2890**
Uncertainty	-0.0021	-0.0066**	-0.0083**	-0.0124***	0.0006	-0.0004	0.0017	0.0008
Leverage	0.0039	0.0183	0.0279	-0.0078	0.0379	-0.1900	-0.2502	-0.1889
MCAP	-0.2557**	-0.3825***	-0.3336*	-0.4719	0.0178	-0.0756	-0.1035	-0.0694
MTB	-0.0029	-0.0085*	-0.0114*	-0.0197***	-0.0398*	-0.0094	0.0093	-0.0298
R^2 *Within*	0.0494	0.0944	0.0710	0.1054	0.2112	0.3129	0.2513	0.3385
	ESG - 10-K/10-Q				**ESG - ESG Reports**			
NPT	-0.0191	0.0275	0.0315	0.0444	0.0305	-0.0652	-0.0296	0.0537
NPT_{t-1}	-0.0478**	-0.0351	-0.0560	-0.1009***	-0.0517	-0.0344	-0.2428**	-0.0824
SP500	0.4560***	0.9342***	0.8213***	0.9454***	0.9268***	1.0911**	0.9956	1.2953**
DFF	0.0103	-0.0022	0.0652	0.0761*	-0.0313	-0.0216	-0.1198**	-0.0050
GDELT	0.0014	0.0022	0.0003	-0.0034	-0.0008	0.0015	-0.0021	-0.0033
GDP	0.4342***	0.6947***	0.7768***	0.3672*	1.1233**	2.4126***	1.9817**	2.6243***
Unrate	0.0001	-0.0002	-0.0003	-0.0021**	0.0007	-0.0004	-0.0021	-0.0018
Spread	0.0531*	0.0458	0.1058*	0.2177***	0.0808	0.2104**	0.2173	0.2908**
Uncertainty	-0.0021	-0.0068**	-0.0084**	-0.0126***	0.0007	-0.0002	0.0012	0.0005
Leverage	0.0036	0.0177	0.0233	-0.0097	0.0492	-0.1808	-0.2134	-0.1621
MCAP	-0.2568**	-0.3724***	-0.3152*	-0.4537	0.0159	-0.0766	-0.1118	-0.0665
MTB	-0.0031	-0.0090*	-0.0137**	-0.0212***	-0.0376*	-0.0077	0.0205	-0.0308
R^2 *Within*	0.0486	0.0938	0.0689	0.1052	0.2187	0.3092	0.2676	0.3414

Note: The table reports panel OLS estimates of CARs over the event windows ([-1,1]), ([-3,3]), ([-5,5]), and ([-1,10]) for 10-K/10-Q filings and ESG reports on the overall and the overall ESG level. All specifications include firm fixed effects and time fixed effects, with two-way clustered standard errors (by firm and time). R^2 *within* is reported. ***, **, and * indicate significance at the 1%, 5%, and 10% levels, respectively.

Table 4.3 reports the pillar-level results for both SEC filings and ESG reports. For annual and quarterly filings, pillar-level ESG sentiment is significantly associated with positive CARs for the environmental and social pillars, while governance sentiment remains insignificant, in line with prior evidence on the limited standalone impact of G (Zheng & Lee, 2026). Environmental sentiment emerges as the most stable driver, showing a positive and robust relationship with REIT returns across three event windows ([-3,3], [-5,5], [-1,10]), whereas social sentiment is weaker and only marginally significant in the [-1,10] window. In the more granular analysis of ESG reports, we similarly find no significant effects for governance-related language, but in this case, environmental sentiment is also insignificant, while the social pillar is associated with

negative CARs in the [-3,3] window. Given that ESG reports are typically framed as CSR or sustainability reports and place strong emphasis on environmental and especially social themes, the social component likely dominates the aggregate sentiment signal in this disclosure channel. The weak or even adverse market reaction to social sentiment in voluntary ESG reports is consistent with the notion that social initiatives are harder to link to immediate, measurable financial outcomes for REITs, and that investors may discount generic or promotional S-language when it is not clearly tied to cash-flow-relevant fundamentals (Carnini Pulino et al., 2022).

Table 4.3 – Regression Results: Pillar-Level ESG Sentiment

	10-K/10-Q vs. ESG reports							
	[-1,1]	**[-3,3]**	**[-5,5]**	**[-1,10]**	**[-1,1]**	**[-3,3]**	**[-5,5]**	**[-1,10]**
	Environmental - 10-K/10-Q				**Environmental - ESG Reports**			
NPT	0.0083	0.0227**	0.0232*	0.0214*	0.0046	0.0020	-0.0060	0.0544
NPT_{t-1}	-0.0092	-0.0132	-0.0132	-0.0210	-0.0577	-0.0820	-0.2054***	-0.0608
SP500	0.4833***	0.9339***	0.8088***	0.9646***	0.8883**	1.1142**	0.9863	1.2710**
DFF	0.0135	-0.0050	0.0684	0.0778*	-0.0443	-0.0271	-0.1439**	-0.0060
GDELT	0.0011	0.0020	-0.0003	-0.0043	-0.0011	0.0008	-0.0026	-0.0035
GDP	0.4322***	0.6411***	0.7544***	0.4112**	1.0900*	2.4100**	1.9053**	2.5898***
Unrate	0.0002	-0.0001	-0.0003	-0.0017	0.0006	-0.0004	-0.0022	-0.0018
Spread	0.0497*	0.0446	0.1157**	0.2093***	0.0812	0.2175**	0.2129	0.2894**
Uncertainty	-0.0028	-0.0081**	-0.0095**	-0.0138***	0.0008	-0.0002	0.0009	0.0004
Leverage	0.0051	0.0126	0.0370	0.0086	0.0482	-0.1559	-0.1941	-0.1673
MCAP	-0.2566**	-0.3579**	-0.2375	-0.3709	0.0158	-0.0725	-0.0966	-0.0582
MTB	-0.0022	-0.0087	-0.0157**	-0.0213***	-0.0381*	-0.0119	0.0045	-0.0378
R^2 *Within*	0.0459	0.0901	0.0662	0.0968	0.2266	0.3170	0.2820	0.3438
	Social - 10-K/10-Q				**Social - ESG Reports**			
NPT	0.0077	0.0208	0.0195	0.0264*	-0.0176	-0.1517*	-0.0625	0.0234
NPT_{t-1}	0.0179	-0.0005	0.0302	0.0265	0.0439	0.0890	-0.0366	-0.0518
SP500	0.5456***	1.1410***	0.9897***	1.1623***	0.8836**	0.9902*	0.9961	1.2974**
DFF	0.0176	0.0016	0.0869	0.0898*	-0.0350	-0.0259	-0.1020**	-0.0035
GDELT	0.0014	0.0044	0.0013	-0.0039	0.0000	0.0029	-0.0003	-0.0032
GDP	0.3257***	0.5207***	0.5993***	0.3571*	0.8951	2.1014**	1.7453*	2.6931***
Unrate	0.0008	-0.0002	-0.0002	-0.0008	0.0009	-0.0003	-0.0018	-0.0019
Spread	0.0371	0.0372	0.1053*	0.2217***	0.0464	0.1682	0.1704	0.2951*
Uncertainty	-0.0032	-0.0102***	-0.0121***	-0.0158***	0.0003	-0.0009	0.0005	0.0005
Leverage	0.0203	0.0386	0.0786	0.0534	0.0355	-0.2379	-0.2679	-0.1762
MCAP	-0.1814	-0.2930*	-0.0977	-0.0879	0.0249	-0.0587	-0.0872	-0.0680
MTB	-0.0043	-0.0165**	-0.0208*	-0.0290**	-0.0453**	-0.0241	-0.0029	-0.0303
R^2 *Within*	0.0437	0.1048	0.0709	0.0997	0.2129	0.3232	0.2163	0.3332
	Governance - 10-K/10-Q				**Governance - ESG Reports**			
NPT	-0.0424	-0.0291	-0.0275	-0.0330	0.0271	0.0306	0.0275	-0.0242
NPT_{t-1}	-0.0262	0.0060	0.0032	-0.0204	-0.0383	0.0099	-0.0340	-0.0871
SP500	0.4498***	0.9342***	0.8221***	0.9462***	0.8837**	1.1688**	1.0942	1.1401
DFF	0.0099	-0.0026	0.0650	0.0763*	-0.0405	-0.0059	-0.0960*	-0.0193
GDELT	0.0012	0.0021	0.0001	-0.0036	-0.0021	0.0008	0.0004	-0.0063
GDP	0.4259***	0.6959***	0.7804***	0.3825*	1.0081*	2.3325**	1.6553*	2.7183***
Unrate	0.0001	-0.0002	-0.0003	-0.0021**	0.0008	-0.0000	-0.0017	-0.0025
Spread	0.0494*	0.0465	0.1072*	0.2192***	0.0574	0.1837	0.1650	0.3055*
Uncertainty	-0.0022	-0.0069**	-0.0085**	-0.0127***	0.0008	-0.0004	0.0004	0.0009
Leverage	0.0061	0.0213	0.0280	-0.0023	0.0023	-0.1921	-0.2560	-0.2187
MCAP	-0.2685**	-0.3953***	-0.3123*	-0.4204	0.0177	-0.0668	-0.0790	-0.0629
MTB	-0.0038	-0.0090	-0.0141**	-0.0228***	-0.0392**	-0.0148	-0.0087	-0.0325
R^2 *Within*	0.0503	0.0950	0.0689	0.1039	0.2329	0.2890	0.2035	0.3396

Note: The table reports panel OLS estimates of CARs over the event windows ([-1,1]), ([-3,3]), ([-5,5]), and ([-1,10]) for 10-K/10-Q filings and ESG reports on the pillar level. All specifications include firm fixed effects and time fixed effects, with two-way clustered standard errors (by firm and time). R^2 *within* is reported. ***, **, and * indicate significance at the 1%, 5%, and 10% levels, respectively.

Since ESG reporting is not mandated by Nareit, the availability of sustainability reports varies across REITs. While the sample is sufficient to analyze sentiment separately for the individual pillars, it does not allow for reliable pillar-level comparisons across asset classes. In contrast, 10-K and 10-Q filings are available for all REITs and therefore support both granular analysis and subsequent sector-level differentiation. Tables 4.4 to 4.7 report the corresponding OLS regression results for the different ESG pillars at the property-type level (industrial, residential, lodging, health care, office, retail, data center, other). The coefficients differ meaningfully in both sign and magnitude, consistent with earlier evidence that tone and ESG and REIT performance in general vary by property type (Carstens & Freybote, 2021; Liang et al., 2021). Environmental sentiment shows a relatively stable and positive association with CARs, with significant effects for several operationally intensive asset classes (industrial, lodging, health care), which is consistent with economic intuition: environmental factors directly affect building performance, operating expenses, and compliance risks. Investors may therefore interpret more positive environmental disclosures as indicative of lower physical or regulatory exposure, particularly when property assets are operationally intensive. Social sentiment also yields positive results in two sectors (lodging, other), although less consistently, which aligns with the limited immediate monetization of social initiatives in real estate valuations. As the measurement of social attributes improves in the overall real estate sector (e.g., tenant well-being indicators, employee-related disclosures), the relevance of S may increase over time. Governance sentiment remains insignificant except for residential and retail REITs. The lack of significance and mixed results in the other subsectors (Kim & Li, 2021; Morri et al., 2024; Zheng & Lee, 2026) are likely due to governance-related risks already being mitigated through REIT-specific regulatory structures, which limits the incremental information contained in narrative governance language.

Overall, the environmental and social pillars of ESG sentiment appear to be more relevant for investors. This is in line with the literature finding that E and S dominate ESG results (Brounen & Marcato, 2018; Fan et al., 2024; Nofsinger et al., 2019) and that they are the most consistent positive drivers (Brounen & Marcato, 2018; Carnini Pulino et al., 2022; Erol et al., 2023; Zheng & Lee, 2026). It supports the argument that aggregated ESG measures obscure critical differences across E, S, and G components and confirms H2, indicating that investors process ESG information in a highly differentiated manner rather than responding to a unified "ESG signal" (Zheng & Lee, 2026).

The subsector analyses in Tables 4.4 to 4.6 further confirm that ESG effects are not uniform across the REIT universe and that value relevance is tightly linked to economic characteristics of the underlying asset classes. Human- and service-intensive sectors such as lodging and health care show the clearest and broadest ESG reactions. The lodging sector exhibits significant positive effects in the [-3,3] and the [-5,5] windows for environmental and social sentiment. The asset

class health care has positive and significant CARs in the same windows for the environmental sentiment as well. The overall ESG results per sector, shown in Table 4.7, also show that the joint ESG sentiment in those asset classes is positively related to CARs in one window each. This is consistent with investors closely monitoring ESG language where day-to-day operations, tenant or patient contact, and reputational risk are particularly present (Liang et al., 2021). At the same time, negative coefficients for lagged ESG sentiment in health care (and other segments with significant lagged effects) suggest that persistent "good ESG news" in these subsectors may eventually reflect information that is already incorporated in prices or raise concerns about potential overstatement.

Energy- and capital-intensive asset classes such as industrial REITs likewise display positive reactions to ESG sentiment, with overall and environmental sentiment showing robust and economically intuitive effects in line with the importance of building efficiency, retrofit needs, and regulatory exposure for these properties. Residential REITs stand out by combining positive overall ESG effects (Table 4.7) with significant governance components in multiple windows, which can be seen in Table 4.6. Here, decisions on rents, maintenance, and tenant treatment directly affect households and attract political scrutiny, so investors may view strong governance language as a credible signal of lower regulatory, reputational, and agency risk than in other property types. By contrast, retail REITs exhibit negative reactions to governance sentiment (Table 4.6), which is consistent with the sector facing ongoing structural challenges where emphatic governance language may be interpreted as downplaying critical times rather than as pure good news. Data center REITs form another distinctive group, with consistently negative and highly significant coefficients across specifications. In a sector with very visible energy and climate concerns, investors appear particularly skeptical toward ESG language, echoing evidence that ESG sentiment is especially salient in industries with pronounced ESG controversies (Jeong et al., 2025).

Together, these patterns show that ESG sentiment is rewarded most strongly in subsectors where ESG issues are closely tied to operating models and that the direction and strength of the effects differ systematically across asset classes rather than reflecting a single, market-wide ESG premium. This asset-class-specific heterogeneity provides empirical support for H3 and underlines why REITs offer a particularly suitable setting for our analysis, as their standardized structures and clearly distinguishable property types allow ESG–return relationships to be studied in a granular, economically meaningful way.

Table 4.4 – Regression Results: Environmental and Sector-Level Sentiment

	Environmental							
	[-1,1]	**[-3,3]**	**[-5,5]**	**[-1,10]**	**[-1,1]**	**[-3,3]**	**[-5,5]**	**[-1,10]**
	Industrial				**Residential**			
NPT	0.0138	0.0210	0.0114	0.0467*	0.0014	0.0147	0.0150	0.0010
NPT_{t-1}	-0.0091	0.0337	0.0116	-0.0146	0.0070	-0.0038	-0.0184	0.0012
SP500	0.4136	0.7499	0.6885	0.2471	0.2818***	0.3086**	0.4014**	0.3009
DFF	0.1216***	0.1696***	0.1750***	0.0804**	0.0176	0.0388	0.1024*	0.1301**
GDELT	-0.0019	-0.0063	-0.0050	-0.0107***	0.0006	-0.0013	-0.0051**	-0.0027
GDP	-0.1795	-0.4740	-0.3149	-0.6268	-0.0925	0.0954	0.3832	-0.0125
Unrate	0.0001	0.0020**	0.0022*	0.0001	0.0001	0.0005	0.0009	-0.0001
Spread	-0.0555	-0.1597***	-0.0945	-0.1055	0.0212	-0.0186	-0.0048	0.0448
Uncertainty	-0.0002	0.0014	-0.0050	-0.0164*	0.0016	0.0025	0.0001	-0.0029
Leverage	-0.1218**	-0.1740***	-0.1856***	-0.3530***	0.0120	0.0048	-0.0576	-0.0463
MCAP	-0.8294***	-1.4885***	-1.0834***	-1.6947***	-0.2489	-0.3600	-0.9621***	-2.3237***
MTB	-0.0042	-0.0021	-0.0089	-0.0152	-0.0297***	-0.0317**	-0.0289**	-0.0465**
R^2 *Within*	0.1636	0.2414	0.1719	0.3008	0.1112	0.1083	0.1500	0.2125
	Lodging				**Health Care**			
NPT	0.0264	0.0995*	0.0917*	0.0610	0.0155	0.0295*	0.0427*	0.0234
NPT_{t-1}	0.0317	0.0200	0.0287	0.0350	-0.0908*	-0.1019	-0.0860	-0.0803
SP500	1.4767**	2.5905***	1.3511	2.1111***	0.3822**	0.6933***	0.8037**	1.1693**
DFF	-0.0198	0.0210	-0.0038	-0.0446	0.0316	0.0024	0.0860	0.0880
GDELT	0.0124	0.0089	0.0113	0.0097	-0.0034	0.0004	-0.0036	-0.0082*
GDP	1.5753***	1.7331***	2.1722***	1.3556***	0.1777	0.1909	0.2338	0.1866
Unrate	-0.0059	-0.0369***	-0.0133	0.0374***	-0.0046	-0.0172***	-0.0218***	-0.0077
Spread	0.2266**	0.4498***	0.6790***	0.5759***	0.0212	-0.0302	0.0204	0.2630**
Uncertainty	-0.0129	-0.0208**	-0.0215**	-0.0283***	-0.0033	-0.0067	-0.0071	-0.0161***
Leverage	0.2246*	0.3381***	0.3602*	0.3435	-0.2117**	-0.2477	-0.2483	-0.2311**
MCAP	-0.1134	0.1799***	0.1712	0.2060	-1.1828*	-2.6320**	-3.3240**	-2.9832***
MTB	0.0094	-0.0412***	0.0015	-0.0053	0.0176**	0.0216	0.0268	0.0200**
R^2 *Within*	0.1438	0.3208	0.1720	0.2379	0.1376	0.2372	0.2364	0.2661
	Office				**Retail**			
NPT	0.0172	0.0180	0.0602	0.0501	0.0101	0.0159	0.0280	0.0241
NPT_{t-1}	-0.0137	0.0036	-0.0621	-0.0391	-0.0032	-0.0044	0.0052	-0.0067
SP500	0.5236***	0.9995***	1.0997**	1.5820***	0.1832	0.7323**	0.6173**	0.9797***
DFF	0.0269	-0.0082	0.1152	0.1295	0.0112	0.0061	0.0698**	0.0617
GDELT	-0.0002	0.0024	-0.0009	-0.0050*	0.0019	0.0042	0.0047	0.0015
GDP	0.5877***	0.9044***	1.3619***	1.1447**	0.3211*	0.6177***	0.6441***	0.3775
Unrate	0.0001	0.0008	0.0008	-0.0034***	0.0029**	0.0012	0.0001	-0.0022
Spread	0.0619*	0.0097	0.0236	0.1474*	-0.0031	0.0106	0.0784	0.2799**
Uncertainty	0.0024	-0.0025	-0.0044	-0.0097*	-0.0032	-0.0094**	-0.0117***	-0.0213***
Leverage	0.0404	-0.0608	-0.1267	-0.2507	0.1290	0.1378	0.2460	0.2157
MCAP	0.6953***	1.1035*	1.8556	2.3237	-0.6763	-0.6694	-0.8747	-1.6159**
MTB	-0.0272**	-0.0348**	-0.0455	-0.0964***	-0.0038	-0.0182	-0.0343**	-0.0074
R^2 *Within*	0.1289	0.1823	0.1821	0.1881	0.0860	0.1414	0.1399	0.1713
	Data Center				**Other**			
NPT	-0.1272***	-0.1083***	-0.0849	-0.1006***	-0.0090	-0.0043	-0.0284	-0.0091
NPT_{t-1}	-0.3276***	-0.3315***	-0.3653	-0.3790	-0.0058	0.0195	-0.0070	-0.0958*
SP500	0.6335***	1.0881***	1.0964	1.3867***	0.4607**	0.7689**	0.7654***	1.0626***
DFF	0.1231***	0.0978	0.2023***	0.2062***	-0.0204	-0.0522***	0.0255	0.0466*
GDELT	-0.0122***	-0.0151**	-0.0130**	-0.0130***	-0.0011	-0.0003	-0.0041	-0.0075**
GDP	0.0443	-0.2602	-0.1148***	-0.1659	0.4439**	0.7853***	0.6293**	0.4834**
Unrate	0.0024***	-0.0015***	-0.0031	0.0084***	0.0003	0.0001	0.0001	-0.0004
Spread	-0.1283***	-0.1667***	-0.1148***	-0.2072***	0.0714	0.0974	0.1414	0.1955*
Uncertainty	0.0052	0.0033	-0.0025	-0.0077	-0.0015	-0.0061**	-0.0085*	-0.0110**
Leverage	-0.6690***	-0.6101	-0.8536***	-1.2468***	-0.0217	0.0595	0.0543	0.0259
MCAP	-6.3578***	-3.6945**	-3.7571***	-10.384***	-0.3992*	-0.1297	-0.4444	-0.9095**
MTB	0.1279**	0.0827**	0.0339	0.2337***	0.0034	-0.0107	-0.0198	-0.0183
R^2 *Within*	0.6380	0.6030	0.4862	0.7028	0.0584	0.1030	0.1072	0.1508

Note: The table reports panel OLS estimates of CARs over the event windows ([-1,1]), ([-3,3]), ([-5,5]), and ([-1,10]) for 10-K/10-Q filings on the environmental level. All specifications include firm fixed effects and time fixed effects, with two-way clustered standard errors (by firm and time). R^2 *within* is reported. ***, **, and * indicate significance at the 1%, 5%, and 10% levels, respectively.

Table 4.5 – Regression Results: Social and Sector-Level Sentiment

	Social							
	[-1,1]	**[-3,3]**	**[-5,5]**	**[-1,10]**	**[-1,1]**	**[-3,3]**	**[-5,5]**	**[-1,10]**
	Industrial				**Residential**			
NPT	0.0090	-0.0075	-0.0206	-0.0073	-0.0132	-0.0082	0.0003	0.0057
NPT_{t-1}	-0.0946**	-0.1130***	-0.0330	-0.0585	-0.0000	-0.0108	-0.0031	0.0069
SP500	0.1815	0.1442	0.1673	0.0621	0.1879**	0.3083**	0.3492	0.1510
DFF	0.1179**	0.1284***	0.1587***	0.1210*	0.0485***	0.0817	0.1795	0.2311*
GDELT	-0.0048	-0.0080	-0.0087	-0.0168***	0.0004	0.0023	-0.0031	-0.0012
GDP	-0.7932*	-1.1721***	-1.1993**	-1.5042***	-0.1751	0.0797	0.3766*	-0.0529
Unrate	0.0012*	0.0034**	0.0039**	0.0021	0.0006	0.0007	0.0012*	0.0005
Spread	-0.0973	-0.2180**	-0.1602	-0.2200**	-0.0031	-0.0293	-0.0164	0.0251
Uncertainty	0.0018	0.0005	-0.0049	-0.0128***	0.0018	0.0012	-0.0024	-0.0043
Leverage	-0.1384	-0.2355	-0.1680	-0.3080*	0.0183	0.0398	-0.0369	-0.0702
MCAP	-1.1700	-2.3574**	-1.5458*	-1.7147***	-0.3547	-0.6312*	-1.1951***	-3.1585***
MTB	0.0146	0.0156	0.0005	0.0160	-0.0283**	-0.0268*	-0.0238	-0.0446
R^2 *Within*	0.2781	0.3789	0.2935	0.3133	0.1432	0.1311	0.1722	0.2646
	Lodging				**Health Care**			
NPT	0.0541	0.1112**	0.0993**	0.1162***	0.0060	-0.0035	0.0094	0.0251
NPT_{t-1}	0.1318	0.0732	0.1753	0.1490	0.0000	-0.0042	0.0034	0.0152
SP500	1.1645*	2.8521***	1.6232	2.1245***	0.4219**	0.7926***	0.8073***	0.9761**
DFF	-0.1099	-0.0738	-0.1382	-0.1675	0.0242	-0.0079	0.1021	0.0997*
GDELT	0.0152	0.0126	0.0184	0.0164	-0.0032	0.0038	-0.0010	-0.0079
GDP	1.1807***	1.2891**	1.3388***	0.9420**	0.1862	0.0803	0.0716	0.1171
Unrate	-0.0092	-0.0407***	-0.0180	0.0336***	-0.0040	-0.0170***	-0.0223***	-0.0061
Spread	0.2011**	0.4193***	0.6028***	0.5930***	-0.0236	-0.0714	-0.0032	0.2063**
Uncertainty	-0.0122	-0.0186***	-0.0214**	-0.0303***	-0.0017	-0.0065	-0.0073*	-0.0150***
Leverage	0.2567	0.5036***	0.6710***	0.5474**	-0.2118***	-0.2725*	-0.2734*	-0.2870**
MCAP	-0.3110	-0.0695	-0.1653	-0.1064	-1.4072**	-2.8641**	-3.4189***	-3.6071***
MTB	0.0121	-0.0318**	0.0130	0.0039	0.0154	0.0165	0.0251	0.0227
R^2 *Within*	0.1471	0.3438	0.2151	0.3046	0.1152	0.2287	0.2231	0.2750
	Office				**Retail**			
NPT	-0.0050	-0.0209	-0.0101	-0.0437	-0.0033	-0.0048	-0.0232	-0.0071
NPT_{t-1}	0.0025	0.0028	0.0240	0.0008	0.0102	0.0044	0.0154	0.0039
SP500	0.4426	1.0823*	1.5113*	2.2608**	0.2847	0.6888**	0.7679***	1.1930***
DFF	0.0491***	0.0264	0.2163***	0.2313	0.0030	0.0057	0.1025**	0.0596
GDELT	-0.0012	0.0030	-0.0020	-0.0066	0.0021	0.0023	0.0007	-0.0011
GDP	0.6580***	0.7696***	1.2197*	1.2629	0.2722	0.6609**	0.6664**	0.3359
Unrate	0.0012**	0.0005	0.0003	-0.0037**	0.0028**	0.0013	-0.0001	-0.0025
Spread	0.0654	-0.0208	-0.0040	0.2001	-0.0022	0.0129	0.0786	0.3186**
Uncertainty	0.0017	-0.0042	-0.0076	-0.0150*	-0.0033	-0.0083*	-0.0106*	-0.0243***
Leverage	0.0053	-0.0526	-0.1790*	-0.3461	0.1665	0.1374	0.3297	0.2617
MCAP	0.8785**	1.6006**	2.9067	3.7295	-0.6521	-0.6187	-0.5620	-1.7144**
MTB	-0.0341*	-0.0494**	-0.0588	-0.1114**	-0.0094	-0.0294	-0.0582**	-0.0089
R^2 *Within*	0.1781	0.2172	0.2245	0.2431	0.0948	0.1465	0.1559	0.1880
	Data Center				**Other**			
NPT	-0.0929***	-0.0767***	-0.1317***	-0.0514***	0.0199*	0.0424***	0.0261	0.0210
NPT_{t-1}	0.0035	-0.0110	0.0639**	-0.0595***	0.0126	0.0169	0.0378	-0.0126
SP500	0.6663***	1.0485**	1.3494**	1.3883***	0.7211**	1.2669***	1.1062**	1.7027***
DFF	0.0969***	0.0846***	0.1831***	0.1587***	-0.0160	-0.0400	-0.0014	0.0699*
GDELT	-0.0100***	-0.0171**	-0.0104	-0.0100***	-0.0031	0.0013	-0.0012	-0.0082*
GDP	-0.3666	-0.6899***	-0.8099***	-0.6741***	0.3137*	0.7361***	0.5246**	0.4260**
Unrate	-0.0033	-0.0059**	-0.0106	0.0044	0.0008	0.0006	0.0007	0.0007
Spread	-0.2326***	-0.2803***	-0.2967***	-0.3308***	0.0454	0.0513	0.0813	0.1523
Uncertainty	0.0006	-0.0029***	-0.0074	-0.0119	0.0012	-0.0069	-0.0113	-0.0082*
Leverage	-0.6961***	-0.4023***	-1.5983	-1.4663***	-0.0116	0.0825	0.0294	0.0824
MCAP	-11.8057***	-8.4912***	-10.2034***	-14.551***	-0.2134	-0.0229	-0.6848	-0.8019
MTB	0.3170***	0.2623***	0.3576***	0.4219***	0.0065	-0.0146	-0.0184	-0.0300
R^2 *Within*	0.7561	0.7310	0.6144	0.7616	0.0916	0.1995	0.1753	0.2171

Note: The table reports panel OLS estimates of CARs over the event windows ([-1,1]), ([-3,3]), ([-5,5]), and ([-1,10]) for 10-K/10-Q filings on the social level. All specifications include firm fixed effects and time fixed effects, with two-way clustered standard errors (by firm and time). R^2 *within* is reported. ***, **, and * indicate significance at the 1%, 5%, and 10% levels, respectively.

Table 4.6 – Regression Results: Governance and Sector-Level Sentiment

	Governance							
	[-1,1]	**[-3,3]**	**[-5,5]**	**[-1,10]**	**[-1,1]**	**[-3,3]**	**[-5,5]**	**[-1,10]**
	Industrial				**Residential**			
NPT	-0.0305	-0.0682	-0.0752	-0.0196	0.0205**	0.0449**	0.0826***	0.0411
NPT_{t-1}	0.0375	0.0632	0.0258	-0.0588	0.0113	0.0101	-0.0203	0.0105
SP500	0.4047	0.7140	0.6533	0.2428	0.2933***	0.4120***	0.5371***	0.4071**
DFF	0.1218***	0.1676***	0.1782***	0.0838**	0.0190	0.0539	0.1210**	0.1383**
GDELT	-0.0022	-0.0065	-0.0057	-0.0102**	0.0001	-0.0005	-0.0030	-0.0019
GDP	-0.1544	-0.4398	-0.2812	-0.5761	-0.1324	0.1308	0.4683*	0.0743
Unrate	0.0001	0.0019**	0.0021*	-0.0000	-0.0003	0.0004	0.0008	-0.0015
Spread	-0.0524	-0.1525***	-0.0900	-0.0913	0.0201	-0.0139	0.0318	0.1019
Uncertainty	0.0002	0.0019	-0.0038	-0.0149*	0.0028*	0.0039	0.0022	-0.0009
Leverage	-0.1278**	-0.2143***	-0.1879***	-0.3548***	-0.0227	-0.0480	-0.1268*	-0.1733**
MCAP	-0.8518***	-1.6199***	-1.1538***	-1.6163***	-0.3149	-0.7521***	-1.3637***	-2.7022***
MTB	-0.0020	0.0058	-0.0010	-0.0114	-0.0187***	-0.0182***	-0.0101	-0.0276*
R^2 *Within*	0.1575	0.2490	0.1685	0.2762	0.1077	0.1085	0.1367	0.2250
	Lodging				**Health Care**			
NPT	-0.2651*	-0.1440	-0.1888	-0.1550	0.0049	0.0300	0.0302	-0.0199
NPT_{t-1}	-0.3009	-0.2214	-0.4941	-0.3592	-0.0202	-0.0066	0.0266	0.0800***
SP500	0.8624	2.3296***	1.3027	1.7207***	0.3490**	0.6362***	0.6848**	0.8757***
DFF	-0.0837	-0.0072	-0.0836	-0.0940	0.0066	-0.0369	0.0436	0.0648
GDELT	0.0106	0.0077	0.0109	0.0104	-0.0020	0.0006	-0.0026	-0.0067
GDP	1.3233***	1.7106***	1.9059***	1.1453**	0.2682	0.3140	0.3149	0.2718
Unrate	-0.0091	-0.0376***	-0.0148	0.0281**	-0.0042	-0.0161***	-0.0204***	-0.0053
Spread	0.2410***	0.4446***	0.5670***	0.5787***	0.0104	-0.0283	0.0066	0.2423**
Uncertainty	-0.0059	-0.0113*	-0.0131*	-0.0198***	-0.0027	-0.0062	-0.0070	-0.0143***
Leverage	0.2815*	0.5397***	0.5249***	0.5053**	-0.1556***	-0.1624	-0.1720*	-0.2334**
MCAP	-0.1736	0.1513	-0.0080	0.1159	-1.1855*	-2.5577**	-2.9556***	-3.1543***
MTB	0.0025	-0.0419***	-0.0022	-0.0145	0.0143**	0.0216	0.0258	0.0218**
R^2 *Within*	0.1392	0.2947	0.1668	0.2501	0.1148	0.2102	0.2048	0.2585
	Office				**Retail**			
NPT	0.0293	0.0273	0.1031	0.1486	-0.0576*	-0.0845***	-0.0797	-0.0758
NPT_{t-1}	-0.0383	-0.0558	0.0409	-0.0414	-0.0509	-0.0358	-0.0352	-0.0678
SP500	0.5542***	0.9920***	1.0732**	1.5604***	0.1356	0.6498**	0.5601*	0.9020***
DFF	0.0267	-0.0072	0.1122	0.1289	0.0103	0.0003	0.0684*	0.0674*
GDELT	-0.0006	0.0022	-0.0015	-0.0055*	0.0015	0.0035	0.0040	0.0007
GDP	0.5543***	0.8868***	1.3543***	1.1427**	0.3116**	0.6965***	0.7519***	0.3835
Unrate	0.0006	0.0007	0.0008	-0.0031***	0.0029**	0.0013	0.0003	-0.0024
Spread	0.0469	0.0044	0.0146	0.1384*	-0.0041	-0.0146	0.0514	0.2818**
Uncertainty	0.0026	-0.0022	-0.0039	-0.0092*	-0.0042	-0.0098**	-0.0123***	-0.0215***
Leverage	0.0412	-0.0482	-0.1180	-0.2469	0.1153	0.1068	0.2003	0.1985
MCAP	0.6801***	1.0683*	1.9081	2.3241	-0.7066	-0.7658	-1.0413*	-1.7091**
MTB	-0.0258**	-0.0315*	-0.0486*	-0.0969***	-0.0045	-0.0185	-0.0326**	-0.0093
R^2 *Within*	0.1373	0.1798	0.1784	0.1869	0.1035	0.1624	0.1550	0.1847
	Data Center				**Other**			
NPT	-0.1318	-0.1121	-0.1580	-0.1557	0.0603	0.0676	0.1438	0.1212
NPT_{t-1}	-0.3118***	-0.3107***	-0.2752***	-0.5885***	0.0232	-0.0303	-0.0253	-0.1716
SP500	0.4984***	0.9559***	0.9475	1.1449***	0.4666**	0.7929***	0.7946***	1.1011***
DFF	0.1109***	0.0874	0.1922***	0.1893***	-0.0257	-0.0523***	0.0143	0.0483
GDELT	-0.0123***	-0.0152**	-0.0134**	-0.0139***	-0.0012	-0.0001	-0.0039	-0.0072**
GDP	-0.2317	-0.5304	-0.3756*	-0.5368**	0.3989**	0.7145***	0.6008***	0.4615**
Unrate	0.0006	-0.0035	-0.0059***	0.0047	0.0004	0.0002	0.0001	-0.0004
Spread	-0.1782***	-0.2152***	-0.1648	-0.2775***	0.0559	0.0798	0.1244	0.1875*
Uncertainty	0.0063	0.0042*	-0.0012	-0.0058	-0.0008	-0.0053*	-0.0071	-0.0092*
Leverage	-0.7025***	-0.6428***	-0.8998	-1.2330***	-0.0162	0.0542	0.0504	0.0369
MCAP	-6.0363***	-3.4430**	-3.5951**	-10.121***	-0.5394**	-0.3198	-0.6078***	-0.8845***
MTB	0.1198***	0.0784***	0.0321	0.2073***	0.0060	-0.0075	-0.0161	-0.0216
R^2 *Within*	0.5986	0.5743	0.4765	0.7205	0.0683	0.1099	0.1125	0.1479

Note: The table reports panel OLS estimates of CARs over the event windows ([-1,1]), ([-3,3]), ([-5,5]), and ([-1,10]) for 10-K/10-Q filings on the governance level. All specifications include firm fixed effects and time fixed effects, with two-way clustered standard errors (by firm and time). R^2 *within* is reported. ***, **, and * indicate significance at the 1%, 5%, and 10% levels, respectively.

Table 4.7 – Regression Results: Overall ESG and Sector-Level Sentiment

	ESG Overall							
	[-1,1]	**[-3,3]**	**[-5,5]**	**[-1,10]**	**[-1,1]**	**[-3,3]**	**[-5,5]**	**[-1,10]**
	Industrial				**Residential**			
NPT	-0.0018	-0.0242	-0.0279	0.0860***	-0.0090	0.0171	0.0392*	0.0094
NPT_{t-1}	0.0520	0.1454	0.1072**	0.0093	0.0064	0.0286	-0.0167	-0.0115
SP500	0.3955	0.7033	0.6194	0.1813	0.2898***	0.4064***	0.5313***	0.4047**
DFF	0.1202***	0.1691***	0.1732***	0.0725*	0.0190	0.0542	0.1227**	0.1393**
GDELT	-0.0013	-0.0055	-0.0038	-0.0092**	0.0002	-0.0004	-0.0030	-0.0019
GDP	-0.1567	-0.4543	-0.2693	-0.6020	-0.1242	0.1387	0.4838**	0.0843
Unrate	0.0001	0.0020**	0.0021*	-0.0000	-0.0002	0.0004	0.0008	-0.0015
Spread	-0.0514	-0.1554***	-0.0866	-0.0972	0.0208	-0.0145	0.0308	0.1020
Uncertainty	0.0001	0.0016	-0.0039	-0.0152*	0.0028*	0.0040	0.0023	-0.0008
Leverage	-0.1251***	-0.1807***	-0.1897***	-0.3521***	-0.0167	-0.0461	-0.1247*	-0.1674**
MCAP	-0.8459***	-1.4009***	-1.0626***	-1.5903***	-0.2940	-0.7063**	-1.3097***	-2.6736***
MTB	-0.0044	-0.0047	-0.0102	-0.0191	-0.0190***	-0.0187***	-0.0107	-0.0280**
R^2 *Within*	0.1595	0.2432	0.1683	0.2897	0.1064	0.1066	0.1324	0.2237
	Lodging				**Health Care**			
NPT	-0.0413	0.1160*	0.0918	0.1208	0.0345	0.0676	0.1202*	0.0718
NPT_{t-1}	-0.1370	-0.2083	-0.3546	-0.2676	-0.2216**	-0.2053*	-0.2132	-0.2118**
SP500	0.8621	2.2896***	1.2843	1.6902***	0.3328*	0.6099***	0.6470**	0.8554***
DFF	-0.0611	0.0024	-0.0650	-0.0785	0.0103	-0.0289	0.0495	0.0659
GDELT	0.0108	0.0082	0.0113	0.0108	-0.0021	0.0004	-0.0029	-0.0069*
GDP	1.3484***	1.6651***	1.8768***	1.1047**	0.2495	0.2967	0.2616	0.1814
Unrate	-0.0084	-0.0370***	-0.0137	0.0290**	-0.0050	-0.0172***	-0.0223***	-0.0077
Spread	0.2492***	0.4285***	0.5559***	0.5638***	0.0206	-0.0243	0.0076	0.2555***
Uncertainty	-0.0070	-0.0122*	-0.0144*	-0.0209***	-0.0029	-0.0064	-0.0072*	-0.0146***
Leverage	0.2599*	0.5206***	0.5032***	0.4773**	-0.1656***	-0.1727	-0.1832	-0.2440**
MCAP	-0.1427	0.1369	-0.0123	0.1012	-1.1469**	-2.5176***	-2.9692***	-3.1908***
MTB	0.0063	-0.0377***	0.0042	-0.0090	0.0119**	0.0189	0.0241	0.0203*
R^2 *Within*	0.1232	0.2932	0.1603	0.2468	0.1368	0.2206	0.2151	0.2665
	Office				**Retail**			
NPT	0.0537	0.0721	0.2102	0.1826	-0.0103	0.0147	0.0154	0.0518
NPT_{t-1}	-0.0597	-0.0211	-0.1429	-0.1282	-0.0398	-0.0303	0.0181	-0.0609
SP500	0.5550***	0.9948***	1.0944**	1.5781***	0.1632	0.6695**	0.5869**	0.9175***
DFF	0.0274	-0.0064	0.1142	0.1299	0.0119	0.0023	0.0705*	0.0691*
GDELT	-0.0004	0.0023	-0.0011	-0.0049	0.0022	0.0043	0.0048	0.0017
GDP	0.5555***	0.8971***	1.3530***	1.1478**	0.3244**	0.6963***	0.7558***	0.3784
Unrate	0.0006	0.0007	0.0007	-0.0031***	0.0028**	0.0014	0.0003	-0.0024
Spread	0.0472	0.0073	0.0164	0.1413*	0.0024	-0.0115	0.0548	0.2825**
Uncertainty	0.0024	-0.0024	-0.0045	-0.0098*	-0.0035	-0.0094**	-0.0119***	-0.0211***
Leverage	0.0399	-0.0584	-0.1153	-0.2522	0.1148	0.1124	0.2135	0.1965
MCAP	0.7058***	1.1356*	1.9455	2.3620	-0.6885	-0.6981	-0.9401	-1.6188**
MTB	-0.0269**	-0.0359**	-0.0508*	-0.0995***	-0.0044	-0.0198*	-0.0347**	-0.0103
R^2 *Within*	0.1404	0.1822	0.1864	0.1903	0.0924	0.1519	0.1482	0.1785
	Data Center				**Other**			
NPT	-0.1612***	-0.1370***	-0.1655	-0.1287**	-0.0516	0.0077	0.0007	0.0101
NPT_{t-1}	-0.9405***	-0.9496***	-1.0057	-1.3728***	0.0531	0.0512	0.0133	-0.1908
SP500	0.5074***	0.9651***	0.9487	1.2263***	0.4724**	0.7970***	0.7971***	1.1024***
DFF	0.1139***	0.0892***	0.1949***	0.1940***	-0.0248	-0.0510***	0.0138	0.0456
GDELT	-0.0131***	-0.0159**	-0.0142*	-0.0141***	-0.0010	0.0000	-0.0035	-0.0068**
GDP	-0.1108	-0.4112	-0.2659	-0.3274*	0.4354**	0.7445***	0.6172**	0.4615**
Unrate	0.0013***	-0.0028***	-0.0051**	0.0070***	0.0003	0.0001	-0.0000	-0.0005
Spread	-0.1644***	-0.2010***	-0.1524	-0.2443***	0.0664	0.0861	0.1273	0.1861*
Uncertainty	0.0062	0.0042	-0.0014	-0.0064	-0.0009	-0.0053*	-0.0073	-0.0094**
Leverage	-0.6539***	-0.5830***	-0.8331	-1.1491***	-0.0187	0.0435	0.0444	0.0387
MCAP	-6.1283**	-3.4724**	-3.6726**	-10.034***	-0.3423	-0.1675	-0.4341	-0.7929**
MTB	0.1059*	0.0608*	0.0144	0.1909***	0.0031	-0.0077	-0.0190	-0.0246
R^2 *Within*	0.6548	0.6188	0.5023	0.7362	0.0606	0.1043	0.1032	0.1467

Note: The table reports panel OLS estimates of CARs over the event windows ([-1,1]), ([-3,3]), ([-5,5]), and ([-1,10]) for 10-K/10-Q filings on the overall ESG level. All specifications include firm fixed effects and time fixed effects, with two-way clustered standard errors (by firm and time). R^2 *within* is reported. ***, **, and * indicate significance at the 1%, 5%, and 10% levels, respectively.

4.6 Conclusion

This study examines how investors respond to ESG-related sentiment in mandatory and voluntary corporate disclosures of U.S. equity REITs. Using RoBERTa as an advanced LLM, we extract ESG pillar-specific sentiment from 10-K and 10-Q filings as well as standalone ESG reports between 2016 and 2025 and relate these measures to CARs around disclosure dates. Our results provide evidence on how investors process ESG narratives across disclosure channels, ESG pillars, and REIT subsectors. Overall, the results suggest that ESG communication in REIT markets is processed selectively rather than as a uniform signal. This highlights the importance of disaggregated approaches when assessing the role of (ESG) disclosures in real estate capital markets.

Overall sentiment within mandatory filings is associated with short-term market reactions. The aggregate ESG level obscures effects that show up clearly when disaggregating the reports by pillar and sector. For the ESG reports, only the social pillar shows negative CAR reactions to the sentiment. Additional information on ESG matters might show overinvesting and potential greenwashing, which is then penalized by investors or simply ignored. In the mandatory filings, environmental and social sentiment exhibit economic relevance, governance-related sentiment is largely insignificant, suggesting limited incremental information content beyond what is already captured by the REIT regulatory framework and standard governance structures. In the pillar-level results, environmental sentiment stands out as the most stable and significant CAR driver, followed by social- and then by governance-related managerial sentiment. The results also reveal substantial heterogeneity across REIT subsectors. Environmental sentiment is particularly relevant for lodging, health care, and residential REITs, where energy efficiency and operating intensity directly affect NOI, while social sentiment plays a more pronounced role in subsectors with high levels of human interaction (e.g., also lodging). Governance sentiment shows mixed results within property types.

From a managerial perspective, the findings suggest that ESG communication strategies should be straightforward and closely aligned with the underlying property portfolio. Simply increasing the positivity of ESG-related language, particularly in voluntary disclosures, does not lead to more favorable market responses and may even cause skepticism in some asset classes. More specific, credible, and economically grounded disclosures are more likely to signal effective risk management and long-term asset quality.

4.7 References

Aroul, R. R., Sabherwal, S., & Villupuram, S. V. (2022). ESG, operational efficiency and operational performance: Evidence from real estate investment trusts. *Managerial Finance*, *48*(8), 1206–1220. https://doi.org/10.1108/MF-12-2021-0593

Baker, S. R., Bloom, N., & Davis, S. J. (2016). Measuring economic policy uncertainty. *The Quarterly Journal of Economics*, *131*(4), 1593–1636. https://doi.org/10.1093/qje/qjw024

Bauer, R., & Hann, D. (2010). *Corporate environmental management and credit risk* [Working paper]. SSRN. https://doi.org/10.2139/ssrn.1660470

Braun, J., Hausler, J., & Schäfers, W. (2020). Artificial intelligence, news sentiment, and property market liquidity. *Journal of Property Investment & Finance*, *38*(4), 309–325. https://doi.org/10.1108/JPIF-08-2019-0100

Brounen, D., & Marcato, G. (2018). *Sustainable insights in public real estate performance: ESG scores and effects in REIT markets* [Working paper]. Lawrence Berkeley National Laboratory. https://bies.lbl.gov/sites/default/files/ESG%20measures%20FinalVersion.pdf

Carnini Pulino, S., Ciaburri, M., Magnanelli, B. S., & Nasta, L. (2022). Does ESG disclosure influence firm performance? *Sustainability*, *14*(13), Article 7595. https://doi.org/10.3390/su14137595

Carstens, R., & Freybote, J. (2021). Can the textual tone in REIT financial statements improve the information environment for commercial real estate investors? An investigation. *Journal of Real Estate Research*, *43*(3), 335–354. https://doi.org/10.1080/08965803.2021.1985923

Cerciello, M., Busato, F., & Taddeo, S. (2023). The effect of sustainable business practices on profitability: Accounting for strategic disclosure. *Corporate Social Responsibility and Environmental Management*, *30*(2), 802–819. https://doi.org/10.1002/csr.2389

Chacon, R. G., Feng, Z., & Wu, Z. (2025). Does investing in ESG pay off? Evidence from REITs. *The Journal of Real Estate Finance and Economics*, *71*, 587–614. https://doi.org/10.1007/s11146-024-09979-y

Chen, Z., & Xie, G. (2022). ESG disclosure and financial performance: Moderating role of ESG investors. *International Review of Financial Analysis*, *83*, Article 102291. https://doi.org/10.1016/j.irfa.2022.102291

Chiang, K. C. H., Wachtel, G. J., & Zhou, X. (2019). Corporate social responsibility and growth opportunity: The case of real estate investment trusts. *Journal of Business Ethics*, *155*, 463–478. https://doi.org/10.1007/s10551-017-3535-1

Cohen, G. (2023). ESG risk and corporate survival. *Environment Systems and Decisions*, *43*, 16–21. https://doi.org/10.1007/s10669-022-09886-8

Contat, J., Hopkins, C., Mejia, L., & Suandi, M. (2024). When climate meets real estate: A survey of the literature. *Real Estate Economics*, *52*(3), 618–659. https://doi.org/10.1111/1540-6229.12489

Devlin, J., Chang, M.-W., Lee, K., & Toutanova, K. (2019). BERT: Pre-training of deep bidirectional transformers for language understanding. In J. Burstein, C. Doran, & T. Solorio (Eds.), *Proceedings of the 2019 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies, Volume 1 (Long and Short Papers)* (pp. 4171–4186). Association for Computational Linguistics. https://doi.org/10.18653/v1/N19-1423

Dhaliwal, D. S., Radhakrishnan, S., Tsang, A., & Yang, Y. G. (2012). Nonfinancial disclosure and analyst forecast accuracy: International evidence on corporate social responsibility disclosure. *The Accounting Review*, *87*, 723–759. https://doi.org/10.2308/accr-10218

Doran, J. S., Peterson, D. R., & Price, S. M. (2012). Earnings conference call content and stock price: The case of REITs. *The Journal of Real Estate Finance and Economics*, *45*(2), 402–434. https://doi.org/10.1007/s11146-010-9266-z

Dorfleitner, G., & Zhang, R. (2024). ESG news sentiment and stock price reactions: A comprehensive investigation via BERT. *Schmalenbach Journal of Business Research*, *76*, 197–244. https://doi.org/10.1007/s41471-024-00185-3

Eichholtz, P., Holtermans, R., Kok, N., & Yönder, E. (2019). Environmental performance and the cost of debt: Evidence from commercial mortgages and REIT bonds. *Journal of Banking & Finance*, *102*, 19–32. https://doi.org/10.1016/j.jbankfin.2019.02.015

Erol, İ., Unal, U., & Coskun, Y. (2023). ESG investing and the financial performance: A panel data analysis of developed REIT markets. *Environmental Science and Pollution Research*, *30*, 85154–85169. https://doi.org/10.1007/s11356-023-28376-1

Fan, K. Y., Shen, J., Hui, E. C. M., & Cheng, L. T. W. (2024). ESG components and equity returns: Evidence from real estate investment trusts. *International Review of Financial Analysis*, *96*(Part B), Article 103716. https://doi.org/10.1016/j.irfa.2024.103716

Fatemi, A., Glaum, M., & Kaiser, S. (2018). ESG performance and firm value: The moderating role of disclosure. *Global Finance Journal*, *38*, 45–64. https://doi.org/10.1016/j.gfj.2017.03.001

Feng, J., Goodell, J. W., & Shen, D. (2022). ESG rating and stock price crash risk: Evidence from China. *Finance Research Letters*, *46*, Article 102476. https://doi.org/10.1016/j.frl.2021.102476

Feng, Z., & Wu, Z. (2023). ESG disclosure, REIT debt financing and firm value. *The Journal of Real Estate Finance and Economics*, *67*, 388–422. https://doi.org/10.1007/s11146-021-09857-x

Gillan, S. L., Koch, A., & Starks, L. T. (2021). Firms and social responsibility: A review of ESG and CSR research in corporate finance. *Journal of Corporate Finance*, *66*, Article 101889. https://doi.org/10.1016/j.jcorpfin.2021.101889

Han, J.-J., Jun, S., & Kim, J. W. (2025). Examining the impact of ESG news sentiment on corporate performance: A comprehensive analysis by news topic and industry. *Business Ethics, the Environment & Responsibility*. Advance online publication. https://doi.org/10.1111/beer.12855

Hausler, J., Ruscheinsky, J., & Lang, M. (2018). News-based sentiment analysis in real estate: A machine learning approach. *Journal of Property Research*, *35*(4), 344–371. https://doi.org/10.1080/09599916.2018.1551923

Huang, J., Hu, P., Wang, D. D., & Wang, Y. (2025). The double signal of ESG reports: Readability, growth, and institutional influence on firm value. *Sustainability*, *17*(6), Article 2514. https://doi.org/10.3390/su17062514

Hunjra, A. I., Mehmood, R., & Tayachi, T. (2020). How do corporate social responsibility and corporate governance affect stock price crash risk? *Journal of Risk and Financial Management*, *13*(2), Article 30. https://doi.org/10.3390/jrfm13020030

Ignatov, K. (2023). When ESG talks: ESG tone of 10-K reports and its significance to stock markets. *International Review of Financial Analysis*, *89*, Article 102745. https://doi.org/10.1016/j.irfa.2023.102745

Jeong, S.-H., Han, J.-J., Jun, S., Kim, S., & Kim, J. W. (2025). Investor responses to ESG news sentiment: Exploring differential effects and industry moderation. *Corporate Social Responsibility and Environmental Management*, *32*(3), 3944–3964. https://doi.org/10.1002/csr.3163

Kakinuma, Y. (2025). ESG in real estate: Sustainability for financial gains or overinvestment by REITs? *Journal of Property Investment & Finance*, *43*(6), 631–648. https://doi.org/10.1108/JPIF-05-2025-0058

Kim, S., & Li, Z. (2021). Understanding the impact of ESG practices in corporate finance. *Sustainability*, *13*(7), Article 3746. https://doi.org/10.3390/su13073746

Koelbl, M. (2020). Is the MD&A of U.S. REITs informative? A textual sentiment study. *Journal of Property Investment & Finance*, *38*(3), 181–201. https://doi.org/10.1108/JPIF-12-2019-0149

K**ohl, N., & Schaefers, W. (2012).** Corporate governance and market valuation of publicly traded real estate companies: Evidence from Europe. *The Journal of Real Estate Finance and Economics*, *44*(3), 362–393. https://doi.org/10.1007/s11146-010-9236-5

Leetaru, K., & Schrodt, P. A. (2025). *GDELT 2.0 event database. GDELT Project*. Retrieved March 14, 2026, from https://www.gdeltproject.org/data.html#documentation

Liang, J., Jain, A., & Wu, H. (2021). Does corporate social responsibility vary by real estate asset types? Evidence from real estate investment trusts. *Sustainability*, *13*(22), Article 12836. https://doi.org/10.3390/su132212836

Liu, N., Zhao, Y., Yan, B., & Hutchison, N. (2025). Tenants' ESG: Influence on preferences and rent premiums for green buildings in commercial real estate. *Journal of Property Research*, *42*(4), 321–355. https://doi.org/10.1080/09599916.2025.2494562

Liu, Y., Ott, M., Goyal, N., Du, J., Joshi, M., Chen, D., Levy, O., Lewis, M., Zettlemoyer, L., & Stoyanov, V. (2019). *RoBERTa: A robustly optimized BERT pretraining approach*. arXiv. https://doi.org/10.48550/arxiv.1907.11692

Malo, P., Sinha, A., Korhonen, P., Wallenius, J., & Takala, P. (2014). Good debt or bad debt: Detecting semantic orientations in economic texts. *Journal of the Association for Information Science and Technology*, *65*(4), 782–796. https://doi.org/10.1002/asi.23062

Morri, G., Anconetani, R., & Benfari, L. (2021). Greenness and financial performance of European REITs. *Journal of European Real Estate Research*, *14*(1), 40–61. https://doi.org/10.1108/JERER-05-2020-0030

Morri, G., Dipierri, A., & Colantoni, F. (2024). ESG dynamics in real estate: Temporal patterns and financial implications for REITs returns. *Journal of European Real Estate Research*, *17*(2), 263–285. https://doi.org/10.1108/JERER-01-2024-0005

Muslu, V., Mutlu, S., Radhakrishnan, S., & Tsang, A. (2019). Corporate social responsibility report narratives and analyst forecast accuracy. *Journal of Business Ethics*, *154*, 1119–1142. https://doi.org/10.1007/s10551-016-3429-7

Nareit. (2025a). *REIT industry fact sheet: Data as of May 30, 2025*. https://www.reit.com/sites/default/files/2025-06/MediaFactSheetMay-2025.pdf

Nareit. (2025b). *REIT industry sustainability report 2025*. https://www.reit.com/investing/reits-sustainability/2025-reit-sustainability-report

Neo, J. R. D., Sing, T. F., & Wang, W. (2025). Doing good and doing well: The relationships between ESG and stock returns of REITs. *Real Estate Economics*, *53*(5), 1105–1137. https://doi.org/10.1111/1540-6229.12534

Nofsinger, J. R., Sulaeman, J., & Varma, A. (2019). Institutional investors and corporate social responsibility. *Journal of Corporate Finance*, *58*, 700–725. https://doi.org/10.1016/j.jcorpfin.2019.07.012

OECD. (2025). *Future-proofing real estate investment: Place-based risks.* OECD Urban Studies. OECD Publishing. https://doi.org/10.1787/2dd12063-en

Paulus, N. M., Lautenschlaeger, L., & Schaefers, W. (2025). Social media and real estate: Do Twitter users predict REIT performance? *Journal of Real Estate Research*, *47*(3), 322–355. https://doi.org/10.1080/08965803.2024.2340915

Robinson, S., & McIntosh, M. G. (2022). A literature review of environmental, social, and governance (ESG) in commercial real estate. *Journal of Real Estate Literature*, *30*(1–2), 54–67. https://doi.org/10.1080/09277544.2022.2106639

Ruscheinsky, J. R., Lang, M., & Schäfers, W. (2018). Real estate media sentiment through textual analysis. *Journal of Property Investment & Finance*, *36*(5), 410–428. https://doi.org/10.1108/JPIF-07-2017-0050

Schiemann, F., & Sakhel, A. (2019). Carbon disclosure, contextual factors, and information asymmetry: The case of physical risk reporting. *European Accounting Review*, *28*, 791–818. https://doi.org/10.1080/09638180.2018.1534600

Schimanski, T., Reding, A., Reding, N., Bingler, J., Kraus, M., & Leippold, M. (2024). Bridging the gap in ESG measurement: Using NLP to quantify environmental, social, and governance communication. *Finance Research Letters*, *61*, Article 104979. https://doi.org/10.1016/j.frl.2024.104979

SEC. (2024). *SEC adopts rules to enhance and standardize climate-related disclosures for investors* (Press Release No. 2024-31). https://www.sec.gov/newsroom/press-releases/2024-31

Tiwari, P., & Miao, J. T. (Eds.). (2022). *A research agenda for real estate*. Edward Elgar Publishing. https://doi.org/10.4337/9781839103933

Tomlinson, B. (2025). *ESG and SEC: Climate disclosure rule update*. EY Insights: ESG Reporting. Retrieved March 14, 2026, from https://www.ey.com/en_us/insights/esg-reporting/esg-and-sec-climate-disclosure-rule-update

Yu, H., Liang, C., Liu, Z., & Wang, H. (2023). News-based ESG sentiment and stock price crash risk. *International Review of Financial Analysis*, *88*, Article 102646. https://doi.org/10.1016/j.irfa.2023.102646

Zheng, Z., & Lee, C. L. (2026). A meta-analysis of ESG factors in the real estate investment trusts sector: Exploring their impacts on REITs performance. *Engineering, Construction and Architectural Management*, *33*(3), 2314–2334. https://doi.org/10.1108/ECAM-05-2024-0652

5 News Sentiment and Stock Return Predictability: The Case of U.S. REITs

5.1 Abstract

The predictive power of investor sentiment in the stock market is commonly attributed to the correction of sentiment-driven mispricing. However, its role in cross-sectional asset allocation remains insufficiently understood. This study investigates whether news-based sentiment exhibits out-of-sample predictive power for weekly subsector returns, using Real Estate Investment Trusts (REITs) as a setting that is particularly prone to such mispricing. We construct weekly sentiment measures from a large corpus of financial and real estate news using two variants of large language models (LLMs). While absolute return prediction remains limited, sentiment contains economically meaningful information for forecasting relative sector performance and improving sector allocation decisions. These findings suggest that sentiment primarily operates as a cross-sectional pricing factor rather than a predictor of aggregate market movements.

Keywords: Financial Forecasting, Real Estate Investment Trusts, Investor Sentiment, Generative Artificial Intelligence, Textual Analysis, Machine Learning, Sector Allocation

5.2 Introduction

Investor sentiment has been recognized as an important driver of asset price dynamics over the past decades. Baker and Wurgler (2007, p. 129) define it as "a belief about future cash flows and investment risks that is not justified by the facts at hand." The concept builds on behavioral models such as that of De Long et al. (1990), in which asset prices may diverge significantly from fundamental values owing to limits to arbitrage arising from the unpredictability of irrational noise traders' beliefs. The influence of investor sentiment on stock returns has been documented across both time series (e.g., W. Y. Lee et al., 2002; Lemmon & Portniaguina, 2006; Tetlock, 2007) and cross-sectional dimensions (e.g., Baker & Wurgler, 2006, 2007; Stambaugh et al., 2012), with predictive power generally attributed to the correction of sentiment-driven mispricing (e.g., Lemmon & Portniaguina, 2006; Stambaugh et al., 2012). Consequently, Baker and Wurgler (2007) argue that the effect of investor sentiment on stock prices is already well documented, shifting the focus toward how to develop an adequate measure and quantify its effects.

Historically, investor sentiment has been proxied using market-based measures (e.g., Baker & Wurgler, 2006; C. M. C. Lee et al., 1991; Neal & Wheatley, 1998; Stambaugh et al., 2012; Swaminathan, 1996) or survey-based indices as more direct measures of sentiment (e.g., G. W. Brown & Cliff, 2004, 2005; Lemmon & Portniaguina, 2006). However, market-based measures may not purely capture investor sentiment but rather reflect the equilibrium outcome of a range of economic forces. Concurrently, surveys often lack high frequency and incentives for truthful responses (Da et al., 2015). Sentiment extraction through textual analysis aims to overcome these limitations. It is based on the expectation that stock market news is processed by investors and can consequently influence their behavior (Tetlock, 2007).[1] However, given the vast amount of textual financial data available today, manual sentiment extraction is increasingly impractical. Recent advances in large language models (LLMs), such as OpenAI's Generative Pre-Trained Transformer (GPT) and Google's Bidirectional Encoder Representations from Transformer (BERT), offer a scalable solution. The alignment of GPT models and trained human annotators is reasonably high across various text-annotation tasks, while GPT models demonstrate higher accuracy than crowd workers at significantly lower cost. This highlights LLMs' potential for efficiency gains in text classification (Churchill et al., 2026; Gilardi et al., 2023).[2]

[1] The predictive power of sentiment extracted from news media content has been demonstrated, for example, by Tetlock (2007) and Tetlock et al. (2008). C. Gu and Kurov (2020) extracted sentiment from Twitter posts, while Renault (2017) analyzed posts from StockTwits. For investor sentiment extraction from corporate disclosures see Price et al. (2012).

[2] Moreover, promising results have been observed in application of GPT models to tasks in the financial domain such as portfolio construction, portfolio optimization, and asset selection by Kim (2023), Ko and Lee (2024), Oehler and Horn (2024), Pelster and Val (2024), and Schneider and Yilmaz (2025).

Although these advances enable scalable and accurate measurement of investor sentiment, a key question is under what conditions such measures provide the most insight into asset pricing dynamics. Baker and Wurgler (2006) suggest that investor sentiment has stronger effects on assets that are hard to value and difficult to arbitrage. Assets that are inherently complex to value are Real Estate Investment Trusts (REITs), as investors must combine information from public equity markets with signals from the illiquid and non-transparent real estate market, where prices are determined by infrequent and backward-looking appraisals (Rochdi & Dietzel, 2015; Ruscheinsky et al., 2018). This may increase valuation uncertainty and limits to arbitrage, thereby amplifying the potential role of sentiment in price formation. Additionally, G. W. Brown and Cliff (2004) reveal that institutional investor sentiment has a stronger impact on the stock market than retail investor sentiment.[3] REIT markets are characterized by high institutional ownership.[4] Therefore, sentiment in these markets likely reflects institutional rather than retail investor behavior. Consistently, institutional investor sentiment exhibits larger effects on REITs than individual investor sentiment (Huerta et al., 2015; Huerta-Sanchez & Escobari, 2018). Despite these characteristics, recent evidence revealed systematic mispricing of risk factors, believed to be largely influenced by investor sentiment (Bond et al., 2025), although error correction from sentiment-induced disequilibrium is more efficient in the REIT market than in common equities (Chiang & Tsai, 2023), consistent with more active reallocation by institutional investors.

Despite extensive evidence on the role of investor sentiment in the stock market, its influence on REIT returns, particularly at high frequency and across subsectors, remains largely unexplored.[5] This gap is notable given the unique structural characteristics of REIT markets, such as informational opacity and institutional dominance, which may amplify the role of sentiment in price formation. This study addresses this gap by examining whether sentiment-driven mispricing in REIT markets can be exploited for return prediction and cross-sectional sector allocation. Using a dataset of more than 320,000 news articles covering the macroeconomy, the REIT market, and the direct real estate market, we construct weekly sentiment measures using both a fine-tuned Robustly Optimized BERT Pre-Training Approach (RoBERTa) model and a GPT-based model. While RoBERTa produces an indicator reflecting aggregate market sentiment, GPT's generalist knowledge is assessed for its ability to capture real estate sector-specific sentiment using a zero-shot prompting approach.[6] We incorporate these sentiment measures into an out-of-sample

[3] Traditionally sentiment effects have been associated with individual investors in small-cap stocks as depicted by C. M. C. Lee et al. (1991), Lemmon and Portniaguina (2006), and Swaminathan (1996).
[4] Bond et al. (2025) state that institutional investors hold, on average, approximately 80% of outstanding REIT shares, substantially exceeding ownership levels in a size- and book-to-market-matched sample of common stocks.
[5] REITs have often been explicitly excluded from empirical analyses. See, for example, Fama and French (1992) and Price et al. (2012).
[6] Zero-shot prompting refers to the approach of using a model without providing examples of correctly executed prompts. Instead, the model relies solely on its pre-existing training, without any form of fine-tuning or additional training.

(OOS) forecasting framework employing both a linear least absolute shrinkage and selection operator (LASSO) model and a nonlinear eXtreme Gradient Boosting (XGBoost) model. The focus is on aggregate return prediction and the relative performance of major U.S. REIT sectors compared to the aggregate index. The prediction of relative sector outperformance is further evaluated in an active sector rotation trading strategy.

Consistent with prior evidence on the difficulty of short-term return forecasting (e.g., Cabrera et al., 2026; Welch & Goyal, 2008), the results reveal limited return predictability at the weekly horizon. However, news-based sentiment contains economically meaningful information for predicting relative sector performance. Particularly, the inclusion of investor sentiment improves the identification of sector outperformance. These signals can be translated into a profitable sector allocation strategy prior to transaction costs, outperforming a non-sentiment benchmark by up to 34% per year. The results remain robust and statistically significant across different trading strategies. Furthermore, the findings indicate a nonlinear influence of investor sentiment on return formation, while all examined news categories are relevant sources of investor sentiment. Moreover, the aggregate REIT market sentiment derived from the fine-tuned RoBERTa model exhibits substantially higher predictive power than the sector-specific sentiment obtained from GPT, thereby challenging the effectiveness of GPT with zero-shot prompting.

We contribute to the literature in several ways. First, we extend research on artificial intelligence in finance by applying different LLMs to REIT-specific sentiment analysis across multiple news sources, thereby highlighting the importance of fine-tuning to properly interpret the REIT market and extract financially relevant sentiment signals. Second, we extend the asset pricing literature by exhibiting that investor sentiment primarily predicts cross-sectional variation in REIT returns. This indicates that sentiment-driven mispricing occurs at the sector level rather than in the aggregate market, reinforcing evidence on the distinct influence of investor sentiment across real estate sectors (e.g., Demiralay & Kilincarslan, 2024; Marcato & Nanda, 2016). Third, we demonstrate the economic relevance of sentiment signals by translating forecasts into a sector rotation strategy, thereby linking return predictability to active asset allocation. Furthermore, to the best of our knowledge, this is the first study to provide a comprehensive analysis of the relationship between weekly sentiment and REIT returns in a cross-sectional OOS analysis.

The remainder of this study is structured as follows. Section 5.3 reviews the related literature and develops the hypotheses. Section 5.4 describes the data, and Section 5.5 outlines the methodology. Section 5.6 presents the empirical results, while Section 5.7 concludes.

5.3 Related Literature and Hypotheses Development

5.3.1 Return Predictability and Relative Sector Performance

The prediction of stock returns has long attracted interest from investors and academics. A large body of literature has assessed different drivers of stock returns, mostly applying linear modeling techniques (e.g., Campbell & Thompson, 2008; Dangl & Halling, 2012; Ferreira & Santa-Clara, 2011; Patelis, 1997; Pesaran & Timmermann, 1994; Rapach et al., 2010). However, Welch and Goyal (2008) demonstrate that the prediction of stock returns is challenging especially after 1975, where neither any common predictive variable alone nor a model incorporating all variables achieved positive out-of-sample R^2 (R^2_{OOS}).[7] In the REIT market, Ling et al. (2000) similarly reported no OOS predictability of REIT excess returns after 1990.

Compared to traditional linear predictive regressions, machine learning models have demonstrated superior OOS performance owing to their ability to capture nonlinear relationships in the data (e.g., Donaldson & Kamstra, 1996; S. Gu et al., 2020; Leow & Lindenthal, 2025). In a large cross-sectional analysis covering all firms listed on the NYSE, AMEX, and NASDAQ, and 94 predictors, S. Gu et al. (2020) demonstrated OOS predictability using tree-based models and neural networks, while unrestricted linear model specifications yielded negative R^2_{OOS}. Leow and Lindenthal (2025) generally confirmed these findings and extended them to the REIT market, exhibiting that sector-specialized REITs (e.g., in the retail, office, or residential sector) were more predictable than those invested in more diversified property types. Conversely, Cabrera et al. (2026) found no OOS predictability for U.S. REITs at the index level using various machine learning models and approximately 125 macroeconomic indicators. Moreover, the predictability of stock returns has been demonstrated to increase with the forecasting horizon (e.g., Ferreira & Santa-Clara, 2011; S. Gu et al., 2020), emphasizing the challenges of return forecasting at high frequencies such as weekly intervals.[8] Consistent with this documented difficulty of high-frequency return prediction and the limited evidence of index-level REIT predictability, we propose our first hypothesis:

H1: *Weekly REIT index returns are not predictable at the aggregate index level.*

While the preceding analysis suggests that return prediction remains elusive, evidence exhibits that investor sentiment primarily affects the cross-section of returns rather than aggregate market movements. Baker and Wurgler (2006) demonstrate that sentiment effects are strongest among stocks that are harder to value and more difficult to arbitrage, implying that sentiment-

[7] The finding of negative values of R^2_{OOS} reported by Welch and Goyal (2008) is not uncommon, as many other studies similarly fail to demonstrate superior performance of traditional predictive regressions relative to the historical mean of stock returns. See, for example, Campbell and Thompson (2008), Ferreira and Santa-Clara (2011), and Rapach et al. (2010).

[8] S. Gu et al. (2020) found R^2_{OOS} to increase by nearly an order of magnitude from monthly to annual forecasts.

driven mispricing varies systematically across assets. Stambaugh et al. (2012) reinforce this finding by depicting that investor sentiment predicts the cross-section of stock returns, with long-short strategies based on sentiment-sensitive anomalies generating significant returns following high-sentiment periods.

This cross-sectional heterogeneity is especially present in real estate, where subsectors present different sensitivity to sentiment and uncertainty (e.g., Carstens & Freybote, 2021; Demiralay & Kilincarslan, 2024; Marcato & Nanda, 2016; Plößl et al., 2023). These findings may indicate that the informational value of sentiment in REIT markets lies not in predicting market movements but in capturing relative shifts in investor expectations across subsectors. If sentiment-driven mispricing varies across sectors, then sentiment signals should be informative for the relative over- or underperformance of individual sectors rather than for aggregated index returns. Nevertheless, the effect of sentiment on stock returns is reversed over the next few trading days (Tetlock, 2007). Measures of news-based investor sentiment are available at sufficiently high frequencies to capture short-term return dynamics, as news articles are continuously published across trading days and weeks. Accordingly, news-based sentiment has depicted predictive power in high-frequency analyses, including intraday, daily, and weekly return predictions (e.g., C. Gu & Kurov, 2020; Renault, 2017; Sun et al., 2016; Zhang et al., 2021). Kirtac and Germano (2024) further demonstrated that a profitable trading strategy can be built solely on news-based investor sentiment, leveraging a fine-tuned LLM. Huerta et al. (2015) support this view by suggesting that investors may use sentiment for capital allocation and portfolio rebalancing decisions. Building on the evidence that sentiment effects may primarily manifest in the cross-section and that REIT subsectors differ meaningfully in their sensitivity to investor sentiment, our second hypothesis is as follows:

H2: *News-based investor sentiment contains economically relevant information for predicting the relative performance of REIT subsectors.*

5.3.2 Large Language Models and Textual Analysis

Prior approaches to textual sentiment analysis in finance mainly relied on dictionary-based approaches (e.g., García, 2013; Tetlock, 2007) and classical machine learning classifiers (e.g., Antweiler & Frank, 2004; Frankel et al., 2022). However, as outlined by Renault (2017), both approaches face considerable limitations: Dictionaries are inherently subjective and include limited word lists, which cannot automatically be adopted to specific domains or text types. This limitation is particularly important for accurate sentiment analysis, as the sentiment of a word may vary depending on its context. Accordingly, domain-specific dictionaries have been found to outperform a general dictionary (Loughran & McDonald, 2011; Price et al., 2012). Conversely, machine learning classifiers overcome this limitation but require sufficiently large pre-labeled

training datasets, which raises reproducibility concerns when training data and model specifications are not disclosed (Renault, 2017).

LLMs mitigate these constraints through different mechanisms depending on their architecture. Encoder-based models such as BERT and its variants (e.g., RoBERTa, FinBERT) are publicly available in their base form and only require limited task-specific fine-tuning rather than a ground-up training process. Generative pre-trained models such as OpenAI's GPT can perform sentiment classification in a zero-shot setting without any training data, relying solely on prompt instructions. Both model families have been applied extensively to financial text classification, while exhibiting superior performance compared to dictionaries and simpler machine learning classifiers (e.g., Amin et al., 2023; Chiu & Hung, 2025; Kirtac & Germano, 2024; Ma et al., 2024; Rathje et al., 2024; Siano & Wysocki, 2021; Suzuki et al., 2025).

Still, the relative performance of these two model families remains inconclusive. While one stream of the literature finds that fine-tuned BERT variants yield superior results (e.g., Rathje et al., 2024; Suzuki et al., 2025), another demonstrates better performance of generative models (e.g., Chiu & Hung, 2025; Fatouros et al., 2023; Kirtac & Germano, 2024). Fatouros et al. (2023) attribute the advantage of generative models to their nuanced contextual understanding, which enables them to capture domain-specific and topic-specific sentiment. Further, they found a GPT model to comprehend the intensity of positive and negative sentiment.

This contextual understanding is especially relevant for the REIT market, where news articles frequently carry heterogeneous implications across subsectors.[9] For example, news regarding changes in industrial production is likely to have a stronger impact on industrial REITs than on residential ones, where the impact may be weaker or even negligible. A generative model prompted to evaluate sector-specific sentiment may thus capture this heterogeneity, whereas an encoder model classifying aggregate article-level sentiment cannot differentiate across sectors by design. Since aggregate market-level sentiment lacks cross-sectional heterogeneity, firm-level investor sentiment demonstrated incremental explanatory power (Mahmoudi et al., 2022). Consistently, we expect sector-level sentiment to have explanatory power beyond market-level sentiment, leading us to our third hypothesis:

H3: *The sector-specific sentiment scores derived from a GPT model carry a stronger predictive signal for REIT returns than the aggregate sentiment derived from a RoBERTa model fine-tuned on general financial text.*

[9] Analyzing the direct U.S. real estate market, Marcato and Nanda (2016) find that the residential and non-residential sector respond differently to investor sentiment. Demiralay and Kilincarslan (2024) demonstrate differences within the sensitivity of various REIT subsectors to uncertainty measures.

5.4 Data

5.4.1 Capital Market and Macroeconomic Data

We use 820 weekly observations from January 2010 to September 2025 for five FTSE EPRA Nareit U.S. total return indices retrieved from LSEG Workspace. The indices included are the Equity REITs, Office REITs, Residential REITs, Industrial REITs, and Retail REITs indices. Additionally, we incorporate (1) financial market data, (2) macroeconomic data, and (3) news articles.

Our weekly returns are computed as the percentage change in index levels from one week's closing value to that of the previous week (*Return*). Momentum is defined as the compounded return over a rolling four-week window (*Return Momentum*), with volatility being measured as the standard deviation of weekly returns over the same rolling four-week window (*Return Volatility*). As a second risk metric, we calculate the Ulcer Index (*Ulcer Index*) for each return series. Additionally, we compute the excess return of each sector against the benchmark sector (*Excess Return*) and its four-week momentum (*Excess Return Momentum*). Additional capital market variables include the default premium, measured as the yield spread between Moody's Baa- and Aaa-rated corporate bonds (*Spread*), and the term premium, defined as the spread between the 10-year Treasury bond yield and the 3-month Treasury bill (*Term*).[10]

Macroeconomic data are obtained from the ALFRED database, including their respective publication dates. Accounting for the publication date enables us to precisely align the information available at each prediction, thereby avoiding any look-ahead bias arising from later revisions or data corrections. We use the unemployment rate (*Unemployment Rate*), the yearly GDP growth (*GDP*), and the yearly CPI growth (*CPI*) as macroeconomic indicators. The descriptive statistics are displayed in Table 5.1.

Table 5.1 – Descriptive Statistics

Metric	Mean	Standard Deviation	Minimum	Median	Maximum
REIT Index Return					
Equity REITs	0.0021	0.0301	-0.2618	0.0032	0.2527
Industrial	0.0028	0.0347	-0.1880	0.0043	0.2201
Office	0.0014	0.0351	-0.2450	0.0038	0.2215
Residential	0.0023	0.0311	-0.2940	0.0034	0.2648
Retail	0.0021	0.0370	-0.3588	0.0039	0.3566
REIT Index Momentum					
Industrial	0.0112	0.0663	-0.3403	0.0149	0.3442
Office	0.0052	0.0643	-0.4014	0.0093	0.2699
Residential	0.0090	0.0524	-0.3992	0.0123	0.2883
Retail	0.0081	0.0642	-0.5234	0.0098	0.3336

To be continued

[10] For the spread between Moody's Baa- and Aaa-rated corporate bonds, we use Thursday values as the end of the week. This results from a one business-day publication lag, implying that Friday values are only available on Monday.

Table 5.1 continued

Metric	Mean	Standard Deviation	Minimum	Median	Maximum
REIT Index Volatility					
Industrial	0.0291	0.0199	0.0029	0.0244	0.1979
Office	0.0288	0.0215	0.0025	0.0232	0.2212
Residential	0.0253	0.0206	0.0016	0.0212	0.2627
Retail	0.0273	0.0268	0.0021	0.0217	0.3223
Sector Ulcer Index					
Industrial	6.2985	4.8724	0.4325	4.9952	24.8823
Office	6.6365	5.5379	0.4778	5.0769	28.9122
Residential	4.9614	4.2072	0.3055	4.0497	26.0373
Retail	6.0316	5.8901	0.1652	4.5755	43.6146
Capital Market Metrics					
Spread	0.0094	0.0024	0.0053	0.0091	0.0191
Term	0.0124	0.0123	-0.0166	0.0143	0.0380
Macroeconomic Metrics					
Unemployment Rate	0.0581	0.0226	0.0340	0.0500	0.1470
GDP	0.0210	0.0229	-0.0954	0.0228	0.1223
CPI	0.0253	0.0236	-0.0746	0.0211	0.1326

Note: This table presents the descriptive statistics of the capital market and macroeconomic variables used in this study. The mean represents the arithmetic average of each variable, while the median corresponds to its 50th percentile. The standard deviation is calculated as the four-week standard deviation of each feature. The minimum and maximum refer to the lowest and highest weekly values observed in the dataset, respectively. For the Equity REITs Index, only the Return statistics are depicted as they primarily serve as a benchmark for sectoral under- and over-performance calculation.

5.4.2 News and Sentiment Data

Since the literature illustrates that both macroeconomic and firm-level news attract investor attention and that their tone consistently affects the stock market (Hirshleifer & Sheng, 2022; Ho et al., 2013; T. Liu & Shi, 2025),[11] and given the high level of institutional ownership in REITs, our dataset comprises three news sources that are primarily followed by institutional investors.[12] These include U.S. real estate news from CoStar, U.S. REIT news, and macroeconomic and policy news,[13] both obtained from S&P Market Intelligence (MI). The rationale for this selection is as follows. Macroeconomic and policy news (*Macro*) provides information about the overall state of the economy as well as the relative attractiveness of the U.S. (REIT) market compared to other international markets. Market-level news (*REIT*) reflects the development of the aggregate market and its subsectors, whereas news on the underlying real estate markets (*Direct RE*) in which REITs operate conveys relevant information about market fundamentals.[14]

CoStar articles include the publication date, headline, and subheading, while articles from S&P Market Intelligence provide the date, headline, and full article text. All three news datasets are

[11] Contrasting evidence is provided by Peng and Xiong (2006), who argue that limited attention causes investors to focus more on market- and sector-level information rather than firm-specific information. Temporary distractions have even been documented for institutional investors, for example, by Kempf et al. (2017).

[12] Most of the existent literature focuses on the textual analysis of a single information source. For firm-level data see Feldman et al. (2010), C. Gu and Kurov (2020), and Price et al. (2012); for market-level data see García (2013), Renault (2017), and Tetlock (2007); for macroeconomic announcements see Boyd et al. (2005) and Gilbert (2011).

[13] A global perspective is chosen owing to the high level of foreign institutional ownership in REITs, which accounts for more than 17% of the total market capitalization of U.S. equity REITs (as of 2024; based on industry data reported by Guevarra and Hudgins (2024)).

[14] In addition, Das et al. (2015) show that sentiment in the direct real estate market influences the capital allocation of institutional investors to the REIT market.

cleaned by removing duplicates and excluding articles without a headline, subheading, or article text. Additionally, market updates are excluded from the CoStar News dataset because these articles are assumed to contain noise rather than true narratives. This results in 50,178 S&P MI REIT News articles, 137,326 S&P MI Macroeconomic and Policy News articles, and 133,720 CoStar News articles. Descriptive statistics for the weekly number of articles are reported in Table 5.2.

Table 5.2 – Descriptive Statistics of Weekly Articles Counts

News Source	Mean	Standard Deviation	Minimum	Median	Maximum
S&P MI REIT News	60.02	30.51	6.00	64.5	166.00
S&P MI Macroeconomic & Policy News	164.27	78.53	21.00	178.00	393.00
CoStar News	159.95	39.40	16.00	159.00	513.00

Note: This table reports the descriptive statistics of the weekly number of available articles from October 24, 2009 to October 31, 2025. The mean represents the average number of articles per week, while the median corresponds to the 50th percentile of the weekly article count. The standard deviation is calculated as the weekly deviation of the weekly article count from the mean. The minimum and the maximum refer to the lowest and highest weekly numbers of articles observed in the dataset, respectively.

For each news source, we compute three sentiment measures, where the prefix corresponds to the news source (*Macro*, *REIT*, or *Direct RE*): First, the weekly sentiment (*Macro Sentiment*, *REIT Sentiment*, and *Direct RE Sentiment*). Second, a smoothed four-week average (*Macro Sentiment Rolling Mean*, *REIT Sentiment Rolling Mean*, and *Direct RE Sentiment Rolling Mean*). Third, we calculate the deviation of the sentiment from its smoothed variant to capture short-term shifts in sentiment that may reflect sudden changes in investor expectations (*Abnormal Macro Sentiment*, *Abnormal REIT Sentiment*, and *Abnormal Direct RE Sentiment*).

5.5 Methodology

5.5.1 Sentiment Classification with LLMs

Building on evidence that RoBERTa and domain-adapted variants outperform the original BERT model (Araci, 2019; Siano & Wysocki, 2021), and that GPT-style transformers achieve higher accuracy (Chiu & Hung, 2025; Fatouros et al., 2023; Kirtac & Germano, 2024), we evaluate recent generations of OpenAI's GPT models (gpt-4.1 and gpt-5) alongside FinBERT and RoBERTa as candidate models for generating sentiment scores used in our return forecasting framework.[15]

[15] The original Transformer architecture, introduced by Vaswani et al. (2017), is a neural network with an encoder and a decoder component. However, modern LLMs, such as BERT or GPT, build only on one of these components. The BERT model is a Transformer encoder, using a bidirectional attention mechanism that processes contextual information from both the left and right of each token (Devlin et al., 2019). Conversely, GPT models use a Transformer decoder, restricting each token to attend solely to preceding tokens. This allows for text generation, where the next output token is chosen to maximize the likelihood of a word conditioned on prior context (Radford et al., 2018). FinBERT differs from the original BERT model in that it is pre-trained on financial text rather than general textual data (Araci, 2019). In comparison, RoBERTa is not only pre-trained on a larger text corpus, but also in an improved setup. Additionally, it was trained exclusively to predict masked tokens and not to predict the next sentence (Y. Liu et al., 2019).

To assess model accuracy in financial sentiment classification, we employ the SEntFiN 1.0 dataset from Sinha et al. (2022), which comprises 10,753 human-annotated financial news headlines and allows direct comparison between model predictions and human labels. All GPT models are tested using an identical prompt instructing the model to classify each news headline as either positive, negative, or neutral (see Appendix I).

Table 5.3 – Performance Metrics for Sentiment Classification on the SEntFiN 1.0 Dataset

Model	Accuracy (%)	Precision (%)	Recall (%)	F1 (%)
gpt-4.1-nano	83.66	83.87	83.71	83.61
gpt-4.1-mini	79.44	80.91	79.61	78.61
gpt-4.1	83.41	83.67	83.58	83.17
gpt-5-nano	77.58	81.74	76.84	77.48
gpt-5-mini	80.50	81.19	80.68	79.96
gpt-5	82.27	82.66	82.47	81.84
FinBERT	73.14	74.16	73.24	73.38
RoBERTa	81.08	81.35	81.15	81.22

Note: This table reports the performance metrics of all evaluated LLMs for sentiment classification on the SEntFiN dataset. All metrics are reported as macro averages as the dataset is approximately balanced across the three sentiment classes. Since the SEntFiN 1.0 dataset includes news headlines, where multiple entities – often with conflicting sentiment – are present, only headlines with one clear sentiment were included in the analysis. This reduces the number of headlines from 10,753 to 7,903. The reasoning parameter for gpt-5-nano and gpt-5-mini has been set to "minimal" and "none" for gpt-5 to improve the reproducibility of the analysis. The RoBERTa model has been fine-tuned on financial text with the Financial PhraseBank dataset (Malo et al., 2014), which is a human-annotated set of 4,840 sentences from financial news.

Since the task involves classifying sentiment in financial news and the RoBERTa model is pre-trained only on general text (Y. Liu et al., 2019), we fine-tune it using the human-annotated Financial PhraseBank dataset from Malo et al. (2014). As reported in Table 5.3, gpt-4.1-nano achieved the highest classification accuracy and is therefore selected as the representative GPT model. The fine-tuned RoBERTa model strongly outperforms the FinBERT model and is thus selected as the comparison benchmark to the GPT model.

5.5.2 Construction of Sentiment Indices

Theoretically, the application of GPT models in sentiment analysis offers the distinct advantage that the model's large training corpus enables it to infer not only the overall sentiment conveyed by an article's language but also the directional impact that the same article may hold for different sectors. Financial news frequently contains information of varying importance across sectors. For example, an article highly relevant to office REITs may have only marginal implications for the retail sector. To account for this heterogeneity, the GPT model is instructed to classify the sentiment of each article separately for each of the five REIT indices in a zero-shot prompting approach, and is thereby expected to generate a more nuanced, sector-specific sentiment indicator.[16] Following Fatouros et al. (2023), who document superior performance of

[16] Employing gpt-4.1-nano in a zero-shot prompting approach assesses its existing knowledge and understanding of the real estate sector. Simultaneously, Churchill et al. (2026) indicate that zero-shot prompts produce results comparable to those obtained from two-shot prompts.

a GPT model under numerical rather than categorical classification schemes, the model is asked to evaluate each article's impact on each REIT index using a numerical score ranging from -1 to 1 in increments of 0.1, representing the most negative to the most positive expected impact, respectively. Negative values imply an anticipated decrease in returns, whereas positive values correspond to an expected increase. Additionally, the model is instructed to label an article as irrelevant for a given index whenever its content is not related to that index (see Appendix II). This approach allows for simultaneous filtering of sector-specific relevance and sentiment classification.[17]

After all articles have been classified, the numerical scores are aggregated into a weekly sentiment indicator by computing the simple weekly average, where a week is defined as running from Saturday through Friday. The sentiment indicators $SI_{GPT,N,R,t}$ in week t are calculated for each news source N and each REIT index R:

$$SI_{GPT,N,R,t} = \frac{\sum_{i=1}^{I} \text{sentiment score}_{i,N,R,t}}{\text{total number of relevant articles}_{N,R,t}} \tag{5.1}$$

Since the fine-tuned RoBERTa model can neither assign nuanced numerical sentiment scores nor differentiate between multiple REIT sectors, all articles are classified categorically. The resulting classifications are then aggregated into a single weekly value for each news source following Antweiler and Frank (2004). Their proposed measure of "bullishness" is defined as the difference between the number of positive and negative news articles divided by the sum of both. Accordingly, we compute the sentiment index $SI_{RoBERTa,N,t}$ for each news source N in week t, which serves as the benchmark against which all GPT-based indices are compared:

$$SI_{RoBERTa,N,t} = \frac{\sum_{i=1}^{I} \text{positive article}_{i,N,t} - \sum_{j=1}^{J} \text{negative article}_{j,N,t}}{\sum_{i=1}^{I} \text{positive article}_{i,N,t} + \sum_{j=1}^{J} \text{negative article}_{j,N,t}} \tag{5.2}$$

5.5.3 Forecasting Framework

For forecasting REIT returns, we adhere to a strict out-of-sample forecasting framework employing an expanding and rolling window. The initial prediction target is the first trading week of 2015, using the preceding 260 weeks (spanning 2010 to 2014) for model training. Following each prediction, the training window expands by one week to incorporate the most recent available information. Once the training data reaches 520 weeks at the beginning of 2020, the approach changes to a rolling window scheme to adapt to changing market environments, while maintaining a substantial history.

[17] The analysis was performed in the OpenAI application programming interface (API) as a zero-shot prompting approach without further fine-tuning. The "temperature" parameter was set to zero and a "seed" (42) has been set to enable reproducibility of results as far as possible.

Since the forecasting framework incorporates sentiment measures derived from LLMs, it is crucial to ensure that a model's sentiment classification performance reflects its true capabilities rather than the exploitation of prior knowledge from the training data, commonly referred to as "look-ahead bias" (Didisheim et al., 2025; Lopez-Lira et al., 2025). Given that GPT models' training corpus is largely sourced from the internet (T. B. Brown et al., 2020), it is plausible that the news articles analyzed in this study and the associated REIT index returns were part of the model's training data. If the model has memorized this data, it may infer an article's publication date and the subsequent REIT index return directly from the article content, rather than deriving sentiment solely from language and general financial or real estate knowledge. This would compromise the validity of the analysis by conflating memorization with genuine analytical reasoning.

To mitigate potential look-ahead bias, we follow the approach proposed by Lopez-Lira et al. (2025) and exclude all potentially memorized articles from the analysis (see Appendix III).[18] After removing these articles, the resulting sentiment scores should reflect the model's general sentiment analysis capabilities rather than any form of data memorization.

Furthermore, in the case that feature scaling is applied, we prevent look-ahead bias by deriving scaling parameters solely from the training fold. These parameters are then applied to the respective validation or test fold. By strictly separating the derivation of scaling rules from the test data, we ensure that no future information leaks into the training process or the final prediction.

5.5.4 Model Description and Evaluation Metrics

Consistent with prior findings on the superiority of machine learning models in capturing nonlinear relationships in the data (e.g., S. Gu et al., 2020; Jenett et al., 2026; Leow & Lindenthal, 2025), we employ two distinct models: XGBoost (Chen & Guestrin, 2016) as a nonlinear model and the LASSO regression (Tibshirani, 1996) as a linear benchmark. This allows us to compare linear and nonlinear specifications and evaluate the role of market sentiment in the prediction of REIT returns. To balance computational resources and model adaptability, the hyperparameters of both models are retuned only every first trading week of each year. This annual retuning not only avoids excessive parameter instability but also reduces the risk of possible overfitting induced by frequent retuning. To ensure a robust retuning process, we apply time series cross-validation. This method preserves the temporal order of observations by validating model

[18] A total of 216 articles from S&P MI REIT News, 1,506 articles from S&P MI Macroeconomic & Policy News, and 400 articles from CoStar News were excluded. Hence, the final sample used for sentiment analysis with gpt-4.1-nano comprises 49,962, 135,820, and 133,320 articles, respectively. The set of articles processed by the GPT model additionally differs slightly from that used for RoBERTa as all date expressions within the articles are masked. This ensures that the model cannot infer an article's publication date, and subsequently, any corresponding date-specific REIT index return.

performance at multiple chronological points in time, thereby leading to a more appropriate calibration of model parameters over the sample period.

The LASSO serves as a regularized regression method that performs variable selection and regularization simultaneously. This can enhance the model's prediction accuracy and interpretability. Formally, the LASSO estimator is defined as the solution to the following minimization problem:

$$\min_{\beta}\left\{\frac{1}{2N}\sum_{i=1}^{N}\left(y_i-\beta_0-\sum_{j=1}^{p}x_{ij}\beta_j\right)^2+\lambda\sum_{j=1}^{p}|\beta_j|\right\} \tag{5.3}$$

Here, λ represents the tuning parameter that controls the strength of the shrinkage. A larger λ results in a more aggressive shrinkage that can therefore set individual parameters to zero. This behavior resembles a feature selection process (Hastie et al., 2009).

For nonlinear modeling, the XGBoost algorithm builds on the gradient boosting framework introduced by Friedman (2001) and serves as a highly efficient and flexible implementation, optimized for both performance and scalability. It constructs an ensemble of *K* regression trees, where the predicted output $\hat{y}_i$ is given by the sum of the predictions from the individual trees f_k. The XGBoost model minimizes the following regularized objective function:

$$L(\phi)=\sum_{i} l(\hat{y}_i, y_i)+\sum_{k}\Omega(f_k) \tag{5.4}$$

where l represents a differentiable convex loss function, measuring the deviation between the prediction $\hat{y}_i$ and the true value y_i. The term $\Omega(f_k)$ penalizes the complexity of the model to prevent overfitting. The function is defined as:

$$\Omega(f)=\gamma T+\frac{1}{2}\lambda\|w\|^2 \tag{5.5}$$

where T is the number of leaves penalized by γ, and λ controls the regularization of the leaf weights w.

To evaluate the models' forecasting performance, we employ multiple metrics. For explicit return prediction, we mainly use the mean absolute error (MAE):

$$MAE=\frac{1}{n}\sum_{i=1}^{n}|y_i-\hat{y}_i| \tag{5.6}$$

and the root mean square error (RMSE):

$$RMSE=\sqrt{\frac{1}{n}\sum_{i=1}^{n}(y_i-\hat{y}_i)^2} \tag{5.7}$$

Additionally, we assess predictive power relative to a naïve benchmark using the R^2_{OOS}:

$$R^2_{OOS} = 1 - \frac{\sum_{(i,t)\in T_3}\left(r_{i,t+1} - \hat{r}_{i,t+1}\right)^2}{\sum_{(i,t)\in T_3} r^2_{i,t+1}} \tag{5.8}$$

Here, R^2_{OOS} denotes the OOS coefficient of determination (S. Gu et al., 2020). This metric evaluates forecasting performance relative to a zero-return benchmark, which is chosen owing to the high noise-to-signal ratio in weekly return data, where estimates of the historical mean return tend to be unstable and may provide a misleading reference point. Furthermore, using a zero-return benchmark avoids mechanically inflating performance measures that may arise when benchmarking against a noisy historical mean.

When predicting return direction or sector outperformance, we switch from a regression-based approach to a binary classification task. Here, we focus on balanced accuracy and the area under the receiver operating characteristic curve (ROC-AUC) metric. Given that financial time series data often exhibits class imbalance, the balanced accuracy serves as a superior metric compared to the standard accuracy.

This metric calculates the arithmetic mean of sensitivity (true positive rate) and specificity (true negative rate), ensuring that the model is neither penalized for minority classes nor rewarded for merely predicting the majority class:

$$Balanced\ Accuracy = \frac{1}{2}\left(\frac{TP}{TP + FN} + \frac{TN}{TN + FP}\right) \tag{5.9}$$

The ROC-AUC quantifies the discriminatory ability of the model. The ROC curve compares the true positive rate (TPR) to the false positive rate (FPR) at various classification thresholds. The AUC score quantifies the probability that the classifier assigns a higher rank to a randomly selected positive instance than to a randomly selected negative one, providing a performance measure independent of the classification threshold. The ROC curve is defined by plotting the TPR as a function of the FPR. Consequently, the AUC corresponds to the integral of this curve over the interval [0, 1]:

$$AUC = \int_0^1 TPR(FPR)d(FPR) \tag{5.10}$$

While the aforementioned classification metrics evaluate the ability to distinguish between classes, investment decisions mainly rely on the relative ranking of available assets.

Thus, for the evaluation of the models' ranking capabilities, we utilize an information coefficient (IC), defined as the Spearman rank correlation between predicted scores and realized outcomes:

$$\rho_s = \mathrm{corr}(\mathrm{rank}(\hat{y}), \mathrm{rank}(y)) \tag{5.11}$$

where the IC measures whether assets with higher predicted scores systematically outperform the set benchmark. This aligns with the evaluation framework based on a cross-sectional allocation perspective, where economic value arises from the ability to correctly rank investments rather than from accurately predicting absolute outcomes.

Finally, to address the "black box" nature of the XGBoost model, we utilize SHapley Additive exPlanations (SHAP) values, which enable us to investigate the importance of individual features.

5.6 Results

5.6.1 Regression Analysis

Table 5.4 presents the results of the return prediction task in a regression setting. It reports that the weekly return-level prediction remains highly challenging across all models, as evidenced by predominantly negative R^2_{OOS} values. The limited predictive performance is consistent across sentiment-based and non-sentiment models, indicating that short-horizon index returns are largely dominated by noise. Besides the R^2_{OOS}, the error metrics (RMSE/MAE) display only marginal differences between the models, suggesting limited incremental forecasting gains at the return level. Further, comparing the results of LASSO and XGBoost, the XGBoost models mostly exhibit lower error metrics, indicating higher performance.

We further assess whether the predicted returns display the correct signs. As reported in Table 5.4, the balanced accuracy mostly hovers close to 0.50 across most configurations. This is in line with the limited explanatory power and exhibits an initial lack of directional predictability. Further, the extremely high share of positive predictions reveals a strong upward bias. This suggests that the model mostly captures the average market movement rather than using predictive information from the features.

While these findings do not indicate model failure, they suggest that weekly REIT index returns provide an inherently low signal-to-noise environment. These findings confirm H1 and are consistent with findings regarding the limited predictability of aggregate REIT returns (e.g., Cabrera et al., 2026; Ling et al., 2000).

Table 5.4 – Results: Return Prediction

Sector	Model	MAE	RMSE	R^2_{OOS}	Balanced. Accuracy	Pred. Pos. Rate
RoBERTa Sentiment						
Global	LASSO	0.0240	0.0367	-0.0184	0.5047	0.8040
Global	XGBoost	0.0237	0.0365	-0.0055	0.4914	0.9563
Industrial	LASSO	0.0241	0.0344	-0.0180	0.5045	0.8161
Industrial	XGBoost	0.0238	0.0341	0.0001	0.4956	0.9571
Office	LASSO	0.0260	0.0383	-0.0309	0.5007	0.8054
Office	XGBoost	0.0259	0.0378	-0.0087	0.4938	0.9589
Residential	LASSO	0.0215	0.0327	-0.0162	0.5114	0.8071
Residential	XGBoost	0.0213	0.0326	-0.0063	0.4896	0.9554
Retail	LASSO	0.0241	0.0408	-0.0093	0.5016	0.7875
Retail	XGBoost	0.0239	0.0408	-0.0061	0.4867	0.9536
GPT Sentiment						
Global	LASSO	0.0240	0.0367	-0.0205	0.5096	0.7924
Global	XGBoost	0.0239	0.0366	-0.0117	0.5067	0.9250
Industrial	LASSO	0.0242	0.0346	-0.0271	0.5078	0.7946
Industrial	XGBoost	0.0241	0.0348	-0.0399	0.5037	0.9286
Office	LASSO	0.0261	0.0383	-0.0322	0.5048	0.7982
Office	XGBoost	0.0259	0.0379	-0.0107	0.5161	0.9214
Residential	LASSO	0.0217	0.0329	-0.0254	0.5154	0.7911
Residential	XGBoost	0.0215	0.0327	-0.0161	0.5085	0.9232
Retail	LASSO	0.0241	0.0407	-0.0028	0.5108	0.7857
Retail	XGBoost	0.0241	0.0404	0.0101	0.4984	0.9268
Non-Sentiment						
Global	LASSO	0.0239	0.0365	-0.0081	0.5108	0.8138
Global	XGBoost	0.0238	0.0366	-0.0156	0.5044	0.8661
Industrial	LASSO	0.0240	0.0343	-0.0090	0.5066	0.8143
Industrial	XGBoost	0.0241	0.0343	-0.0108	0.5008	0.8875
Office	LASSO	0.0259	0.0381	-0.0212	0.5124	0.8179
Office	XGBoost	0.0259	0.0379	-0.0119	0.5150	0.8857
Residential	LASSO	0.0215	0.0326	-0.0069	0.5115	0.8196
Residential	XGBoost	0.0214	0.0329	-0.0248	0.5019	0.8375
Retail	LASSO	0.0240	0.0406	0.0033	0.5126	0.8036
Retail	XGBoost	0.0239	0.0410	-0.0162	0.5004	0.8536

Note: This table presents the results of the regression-based return forecasting. The metrics presented for model evaluation are the MAE, RMSE, and R^2_{OOS}. Following S. Gu et al. (2020), R^2_{OOS} is calculated against a zero mean instead of a historical mean, owing to the high level of noise in the data. Additionally, the Balanced Accuracy and the share of positive results predicted is reported.

5.6.2 Directional Movement Prediction

For further testing of market timing prediction, we framed the problem as a pure classification rather than a regression task. The results are presented in Table 5.5. The model performance is mainly assessed using balanced accuracy, capturing the directional classification quality under the observed class imbalance, and ROC-AUC, which reflects the model's ability to rank positive versus negative outcomes independent of a fixed threshold. Confirming the patterns found in the regression output, the balanced accuracy is close to 0.50 for most specifications, indicating a near-random classification performance. A modest exception from this emerges for the residential sector, where the LASSO model with RoBERTa sentiment achieves a score of 0.5346 and a ROC-AUC of 0.5357, suggesting limited but non-negligible predictive structure in this segment. However, the general pattern of ROC-AUC values provides no consistent evidence of a robust ranking signal.

Confirming the regression results, XGBoost classifications exhibit severe bullish bias, effectively collapsing into an "always-long" classifier that often yields a Predicted Positive Rate of over 0.98. This is less dramatically present for the LASSO benchmark, although the overall improvement is limited. Rather than indicating random behavior, the results again suggest that especially the XGBoost model mainly learns the unconditional upward market drift often observed in equity markets. This behavior might be further promoted by the observed slight class imbalance that favors positive predictions.[19]

Overall, the previous observations in return-level and directional predictions suggest that sentiment may not provide sufficient information for absolute market timing at the index level at a weekly horizon. Having confirmed the limited aggregate predictability of returns stated in H1, we turn to testing H2 by examining whether sentiment carries informational value for relative subsector performance.

Table 5.5 – Results: Directional Return Prediction

Sector	Model	Balanced Accuracy	ROC-AUC	Pred. Pos. Rate	TP	FP	TN	FN
RoBERTa Sentiment								
Global	LASSO	0.5092	0.5051	0.7647	963	750	244	283
Global	XGBoost	0.4932	0.4704	0.9482	1174	950	44	72
Industrial	LASSO	0.5043	0.4977	0.7696	249	182	56	73
Industrial	XGBoost	0.4894	0.4653	0.9500	303	229	9	19
Office	LASSO	0.4924	0.4983	0.7429	220	196	65	79
Office	XGBoost	0.4891	0.4744	0.9500	281	251	10	18
Residential	LASSO	0.5346	0.5357	0.7964	262	184	59	55
Residential	XGBoost	0.4984	0.5004	0.9446	299	230	13	18
Retail	LASSO	0.5036	0.4861	0.7500	232	188	64	76
Retail	XGBoost	0.4962	0.4410	0.9482	291	240	12	17
GPT Sentiment								
Global	LASSO	0.5001	0.4896	0.7808	973	776	218	273
Global	XGBoost	0.4990	0.4581	0.9848	1226	980	14	20
Industrial	LASSO	0.5000	0.4836	0.7857	253	187	51	69
Industrial	XGBoost	0.4949	0.4793	0.9857	316	236	2	6
Office	LASSO	0.4831	0.4802	0.7482	219	200	61	80
Office	XGBoost	0.5046	0.4452	0.9857	296	256	5	3
Residential	LASSO	0.5177	0.5024	0.8143	263	193	50	54
Residential	XGBoost	0.5003	0.4783	0.9839	312	239	4	5
Retail	LASSO	0.4975	0.4887	0.7750	238	196	56	70
Retail	XGBoost	0.4962	0.4301	0.9839	302	249	3	6

To be continued

[19] As the pronounced "always-long" behavior may reflect a classification-threshold artifact rather than an inherent model property, we examined whether alternative decision thresholds would mitigate the bias and improve overall model performance. However, the tested adjustments did not improve the results in an economically or statistically meaningful way. Consequently, we identify the weak directional prediction performance as a model characteristic and did not pursue this approach further. These results reinforce the notion that simple directional classification is not the appropriate framework for extracting economic value from sentiment signals.

Table 5.5 continued

Sector	Model	Balanced Accuracy	ROC-AUC	Pred. Pos. Rate	TP	FP	TN	FN
Non-Sentiment								
Global	LASSO	0.5035	0.4933	0.8500	1063	841	153	183
Global	XGBoost	0.4951	0.4655	0.9835	1220	983	11	26
Industrial	LASSO	0.5016	0.4907	0.8589	277	204	34	45
Industrial	XGBoost	0.4949	0.4832	0.9857	316	236	2	6
Office	LASSO	0.5162	0.4966	0.8143	248	208	53	51
Office	XGBoost	0.4921	0.4591	0.9839	292	259	2	7
Residential	LASSO	0.5065	0.5058	0.8839	282	213	30	35
Residential	XGBoost	0.4947	0.4671	0.9857	311	241	2	6
Retail	LASSO	0.4870	0.4733	0.8429	256	216	36	52
Retail	XGBoost	0.4986	0.4526	0.9786	301	247	5	7

Note: This table presents the results of the return direction forecasting. The metrics presented for model evaluation are the Balanced Accuracy, ROC-AUC, and Predicted Positive Rate. TP, FP, TN, and FN resemble the confusion matrix of the models.

5.6.3 Directional Excess Return Prediction

We define the excess return prediction as the outperformance of a sector index versus the overall Equity REITs index. As illustrated in Table 5.6, the models demonstrate moderate but consistent predictive skills across specifications. Globally, the XGBoost model incorporating RoBERTa sentiment performs best, achieving a balanced accuracy of 0.5192. LASSO performs best for both the GPT-Sentiment and the Non-Sentiment models, with values of 0.5152 and 0.5223, respectively. These magnitudes indicate moderate predictive power, reflecting that the excess return framing stabilizes classification outcomes relative to directional return prediction.

At the sector level, a clear cross-sectional pattern is visible. Both balanced accuracy and ROC-AUC consistently indicate stronger predictability for the office and industrial sectors, with values exceeding 0.53 for the best-performing models. Conversely, the residential and retail sectors remain closer to the random benchmark, reflecting limited cross-sectional structure. This pattern is consistent across model specifications and evaluation metrics.

The stronger predictability in the office and industrial sectors is likely to reflect their greater sensitivity to macroeconomic conditions and capital market dynamics, whereas the residential and retail sectors may be more closely linked to household consumption and exhibit weaker responsiveness to investor sentiment. Notably, improved performance is not driven by a one-sided classification behavior, as the confusion matrices reported in Table 5.6 indicate a substantially more balanced classification compared to those from the directional prediction task reported in Table 5.5.

Table 5.6 – Results: Directional Excess Return Prediction

Sector	Model	Balanced Accuracy	ROC-AUC	TP	FP	TN	FN
RoBERTa Sentiment							
Global	LASSO	0.5094	0.5153	568	549	573	550
Global	XGBoost	0.5192	0.5305	557	516	606	561
Industrial	LASSO	0.5177	0.4948	212	172	86	90
Industrial	XGBoost	0.5383	0.5336	180	134	124	122
Office	LASSO	0.4938	0.5123	67	81	219	193
Office	XGBoost	0.5336	0.5776	118	116	184	142
Residential	LASSO	0.5063	0.5093	219	192	72	77
Residential	XGBoost	0.4910	0.4968	155	143	121	141
Retail	LASSO	0.4613	0.4715	70	104	196	190
Retail	XGBoost	0.4950	0.5024	104	123	177	156
GPT Sentiment							
Global	LASSO	0.5152	0.5131	561	529	593	557
Global	XGBoost	0.5095	0.5067	635	616	506	483
Industrial	LASSO	0.4859	0.4928	194	173	85	108
Industrial	XGBoost	0.5075	0.5030	193	161	97	109
Office	LASSO	0.5353	0.5291	79	70	230	181
Office	XGBoost	0.5255	0.5304	145	152	148	115
Residential	LASSO	0.5020	0.4840	212	188	76	84
Residential	XGBoost	0.5123	0.4989	171	146	118	125
Retail	LASSO	0.4828	0.4785	76	98	202	184
Retail	XGBoost	0.4806	0.4759	126	157	143	134
Non-Sentiment							
Global	LASSO	0.5223	0.5164	569	521	601	549
Global	XGBoost	0.5111	0.5242	557	534	588	561
Industrial	LASSO	0.4858	0.4693	215	191	67	87
Industrial	XGBoost	0.5021	0.5059	171	145	113	131
Office	LASSO	0.5155	0.5172	54	53	247	206
Office	XGBoost	0.5255	0.5605	119	122	178	141
Residential	LASSO	0.5083	0.5007	237	207	57	59
Residential	XGBoost	0.5087	0.5100	161	139	125	135
Retail	LASSO	0.5045	0.4828	63	70	230	197
Retail	XGBoost	0.4905	0.4997	106	128	172	154

Note: This table presents the results of the excess return directional forecasts. The metrics presented for model evaluation are the Balanced Accuracy and the ROC-AUC. TP, FP, TN, and FN resemble the confusion matrix of the models.

Comparing sentiment-enhanced models to the non-sentiment baseline demonstrates that the incremental value of sentiment lies primarily in cross-sectional differentiation. For instance, in the office sector, the RoBERTa-Sentiment XGBoost model achieves a ROC-AUC of 0.5776 compared to 0.5605 for the non-sentiment baseline. Similar improvements are observed for the industrial sector, which also persist at the global level, where the sentiment-enhanced RoBERTa XGBoost model clearly outperforms all other specifications. This finding highlights the role of investor sentiment in identifying relative sector outperformance, thereby confirming H2.

The difference between the LASSO benchmark and the XGBoost models offers additional insights into the nature of the sentiment signals. For the RoBERTa sentiment, the LASSO model fails to fully utilize the sentiment signals. This suggests that the informational content of the RoBERTa sentiment unfolds in a regime-dependent or nonlinear way. However, XGBoost exploits such nonlinearities and interaction structures, effectively translating the signals into superior performance. This is consistent with prior findings on the asymmetric effect of investor sentiment

in general and institutional investor sentiment in particular (e.g., Chiang & Tsai, 2023; Huerta-Sanchez & Escobari, 2018).

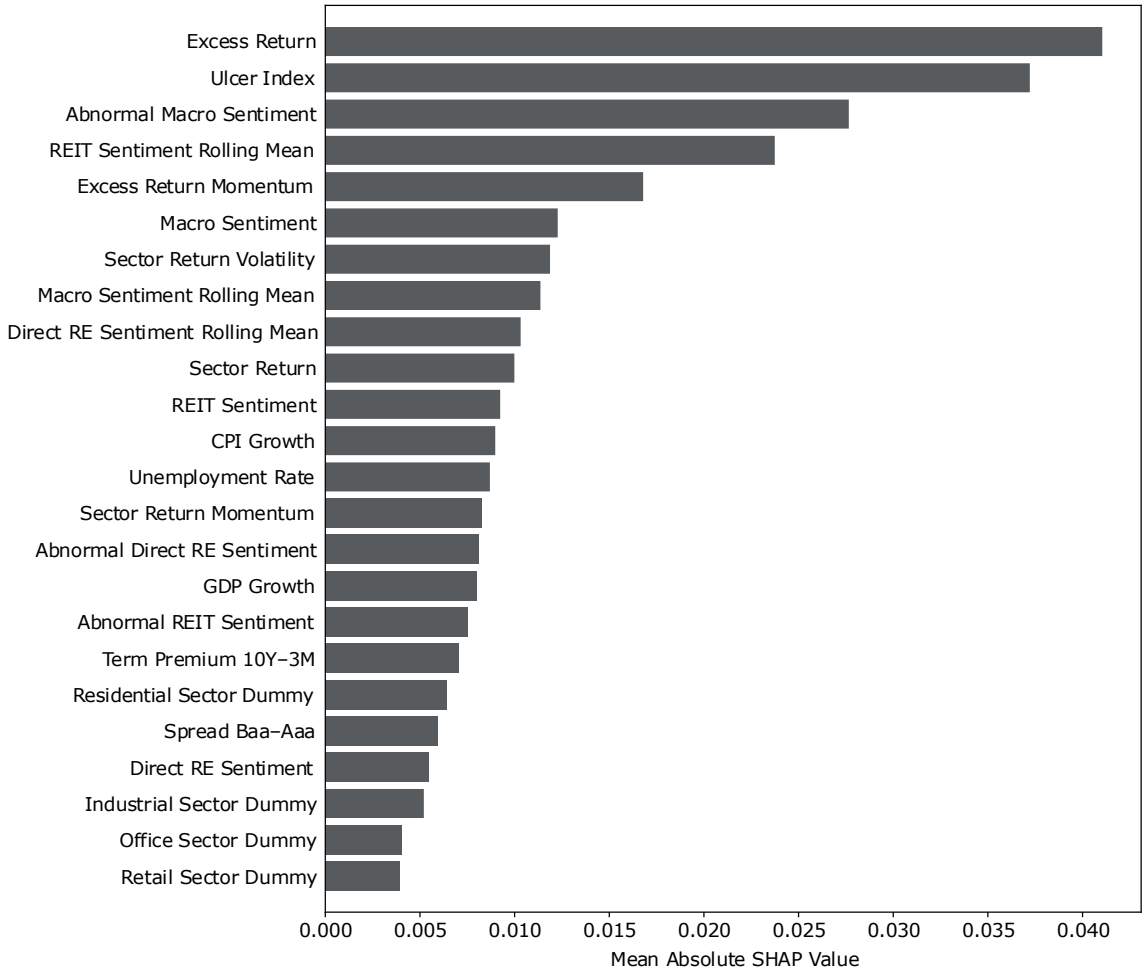
For the GPT-Sentiment model and the non-sentiment baseline, the LASSO occasionally – and at a global level – outperforms XGBoost. This provides insights into the underlying signal quality. When low-information or high-noise signals are present, the regularization of the LASSO acts as an effective filter, preventing overfitting. This appears to be the case for the GPT sentiment. Since the XGBoost is less strictly regularized, it tends to overfit to the noisy GPT-derived signals, leading to weaker OOS performance. While this underscores that complex models such as XGBoost mostly add economic value when fed with high-quality predictive features, it offers two important implications for the GPT sentiment. First, its underperformance relative to both the RoBERTa-Sentiment models and the Non-Sentiment models raises the question of whether sector-specific sentiment measures are less informative than aggregate market sentiment measures. This would contrast findings by Mahmoudi et al. (2022), who document the predictive relevance of cross-sectional sentiment owing to its heterogeneity. Second, it questions the effectiveness of GPT models for sentiment analysis in complex, real estate-specific settings, thereby providing further evidence that GPT models face limitations in tasks involving implicit cues, higher cognitive abilities, and domain-specific knowledge as outlined by Churchill et al. (2026). While this does not imply a general inferiority of generative LLMs, it emphasizes that zero-shot prompting in models such as GPT may be insufficient to capture the nuanced, cross-sectional ranking signals required for quantitative asset allocation in a real estate context. Simultaneously, this finding contrasts with the broader literature on applications of generative LLMs in finance, which often reports strong model performance even without additional fine-tuning (e.g., Fatouros et al., 2023; T. Liu & Shi, 2025). Consequently, we reject H3 and instead highlight the importance of task-specific fine-tuning of LLMs. This may include market-specific characteristics, as the training data of such models may be less comprehensive for certain industries compared to the broader economy or stock market.

For a deeper investigation of driving factors for model performance, we calculated SHAP values for the RoBERTa-Sentiment XGBoost model. Figure 5.1 displays the mean absolute SHAP values, which reflect the contribution of each predictor to the directional excess return prediction.

The most dominant features are past performance and risk metrics, with the *Excess Return* and the *Ulcer Index* suggesting that performance expectations, volatility regimes, and risk-aversion are important drivers of cross-sectional sector dispersion. Crucially, this analysis does validate the importance of soft information, as *Abnormal Macro Sentiment* ranks as the third most important feature, exceeding the contribution of traditional hard macroeconomic indicators such as *GDP* and *CPI*. This might be because of the time lag in the publication process of official metrics,

indicating that macroeconomic sentiment might reflect the current economic expectations of investors. Furthermore, other sentiment indicators such as the *REIT Sentiment Rolling Mean*, the *Macro Sentiment Rolling Mean*, the *Direct RE Sentiment Rolling Mean*, and the *REIT Sentiment* comprise the top half of features. This underlines that narrative-driven market perception plays a pivotal role in predicting relative sector performance.

Figure 5.1 – Out-of-Sample SHAP Absolute Feature Importance



Moreover, the particularly strong impact of macroeconomic news sentiment on returns, measured by the contribution of the respective features to the model's predictions, is consistent with evidence from T. Liu and Shi (2025), who document higher importance of macroeconomic sentiment relative to firm-specific sentiment for common stocks. The high importance of macroeconomic and policy news in the U.S. REIT market may be attributed to its importance for international institutional investors, as well as its broader implications for both the REIT market and the underlying direct real estate market. Consistent with theoretical expectations, news about the real estate market in which REITs invest appears to have a less immediate impact on returns compared to information directly related to the REIT market.

The explicit sector dummy variables exhibit low feature importance. This implies that the model relies on time-varying factors such as sentiment and risk metrics to differentiate between sectors, rather than sector fixed effects. Overall, the presented results of the excess return direction prediction task indicate that market sentiment provides incremental but important information beyond non-sentiment predictors, particularly when framed as a relative performance signal.

5.6.4 Long Signal Trading Strategy

Each week, we allocate the invested capital to the sector with the highest outperformance probability.[20] The results presented below reflect gross returns excluding transaction costs. Table 5.7 displays the results of the proposed trading strategy for all models.

While the LASSO models are included throughout the empirical analysis as a linear benchmark model, the trading strategy is only implemented for the XGBoost model. This choice reflects the economic objective of the proposed trading strategy. The portfolio construction relies on cross-sectional ranking and sector-level differentiation instead of pooled model classification performance. Across the reported specifications, the XGBoost demonstrates stronger subsector-level predictive performance and higher-ranking ability as reflected in sector-level metrics. These are more relevant for generating tradable signals that underlie the proposed strategy.

Table 5.7 – Results: Top-1 Long Strategy

Sector	CAGR	CAGR vs. Sector	Annual Volatility	Sharpe Ratio	IC Mean	IC t-stat	Avg. Turnover
RoBERTa Sentiment							
Top-1 Long	0.0885		0.2451	0.4718	0.0777	2.9906	0.5795
Equity REITs	0.0487	0.0398	0.2278	0.3235			
Industrial	0.1139	-0.0254	0.2455	0.5619			
Office	0.0009	0.0876	0.2719	0.1393			
Residential	0.0596	0.0289	0.2341	0.3667			
Retail	0.0271	0.0614	0.2932	0.2393			
GPT Sentiment							
Top-1 Long	0.0619		0.2400	0.3729	0.0054	0.2171	0.6241
Equity REITs	0.0487	0.0132	0.2278	0.3235			
Industrial	0.1139	-0.0520	0.2455	0.5619			
Office	0.0009	0.0610	0.2719	0.1393			
Residential	0.0596	0.0023	0.2341	0.3667			
Retail	0.0271	0.0348	0.2932	0.2393			
Non-Sentiment							
Top-1 Long	0.0720		0.2451	0.4092	0.0636	2.4723	0.5348
Equity REITs	0.0487	0.0233	0.2278	0.3235			
Industrial	0.1139	-0.0419	0.2455	0.5619			
Office	0.0009	0.0711	0.2719	0.1393			
Residential	0.0596	0.0124	0.2341	0.3667			
Retail	0.0271	0.0449	0.2932	0.2393			

Notes: This table presents the results of the Top-1 Long Strategy. The metrics presented for model evaluation are CAGR, Annual Volatility, Sharpe Ratio, and IC Mean. The CAGR vs. Sector displays the relative outperformance of the proposed strategy versus the benchmark and the individual sectors. The IC t-stat exhibits the statistical significance of the IC Mean. Further, the Avg. Turnover resembles the average frequency of capital reallocation per week.

[20] This represents a rudimentary allocation exercise instead of a fully optimized, production-ready portfolio strategy.

Results are primarily evaluated using the Compound Annual Growth Rate (CAGR). The strategy based on the RoBERTa-Sentiment model achieves the highest economic performance with approximately 398 bps improvement over the Equity REITs index, which serves as a benchmark. For this model, IC statistics indicate a statistically significant and stable ranking signal (IC t-stat > 2), suggesting persistent cross-sectional predictive power. The Non-Sentiment model already depicts modest performance, indicating that some gains over the index performance stem from structural regime and overall capital market dynamics. The GPT-Sentiment model does not outperform the non-sentiment benchmark, contradicting the expectation of increased predictive power through sector-specific sentiment measures as stated in H3.

The weekly turnover rate hovers between 0.53 and 0.63, indicating realistic but non-negligible transaction cost sensitivity. Based on these turnover rates, the break-even transaction costs for the best-performing model are approximately 12.4 bps (one-way). Consequently, and consistent with prior findings from both the stock and the REIT market, excess returns potentially diminish after accounting for transaction costs (e.g., Brooks & Tsolacos, 2003; Ling et al., 2000; Mei & Gao, 1995; Qi & Maddala, 1999). Tetlock (2007) suggests that this may limit investors in high-frequency arbitrage activities, thereby preventing markets from efficiently responding to information from news.

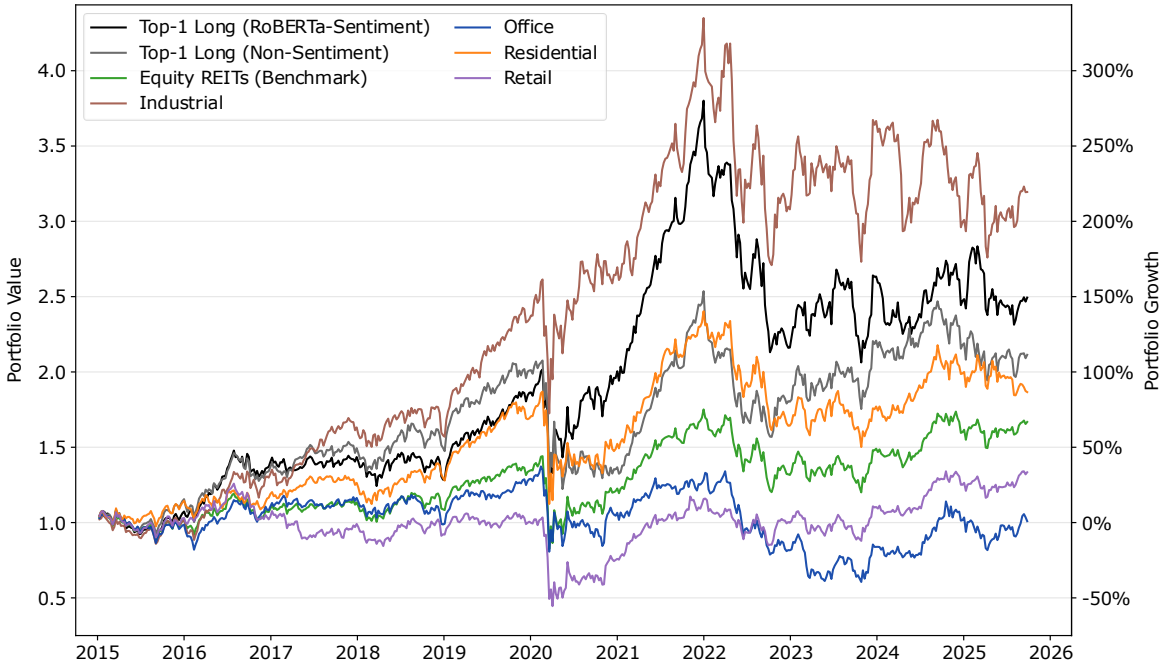
Turning to the strategy's volatility, the Top-1 Long Strategy presents higher levels than the proposed benchmark. However, the increase in the Sharpe ratio shows that the return gains overcompensate for higher risk. Further, we note that the proposed trading strategy could not beat the Industrial REITs index. While the industrial sector exhibits the strongest absolute performance ex post, the strategy identifies leading sectors ex ante under uncertainty rather than replicating a single dominant exposure. Figure 5.2 displays the overall investment performance over time of the proposed RoBERTa-Sentiment model, the Non-Sentiment model, and the individual sectors.

The consistency between classification metrics and realized trading performance strengthens the interpretation that sentiment signals are economically meaningful in an OOS analysis. The improvement in incremental performance relative to the non-sentiment model supports the conclusion that sentiment contributes additional information rather than merely proxying market fundamentals.

Additionally, the observed discrepancy between moderate classification metrics and stronger trading performance suggests that the models capture relative ranking information rather than precise probabilistic classification ability. While average classification performance remains limited, correct ranking at the top of the distribution, as implemented with the Top-1 allocation rule used in the trading strategy, appears sufficient to generate economically meaningful

allocation decisions. This highlights that allocation-based evaluation frameworks may be more informative than pure classification metrics in cross-sectional sector rotation settings, as economic value primarily arises from correctly identifying the top-ranked asset rather than achieving uniformly high classification accuracy. This behavior is consistent with a ranking-based allocation mechanism, where economic value arises from correctly identifying leading sectors during sustained trend episodes rather than from uniformly accurate week-to-week classification.

Figure 5.2 – Top-1 Long Trading Strategy



Furthermore, as a robustness check, we extended the analysis to a Top-2 Long Strategy, each week investing equally in the two sectors with the highest outperformance probability. This allows us to examine whether the model's ranking capability remains informative beyond the single top-ranked sector. If the ranking truly captures the relative performance across sectors, we would expect additional added value rather than a deterioration in strategy performance. Table 5.8 exhibits that, when comparing both trading strategies, an encouraging pattern emerges for the RoBERTa-Sentiment model. Under the split allocation decision, we observe an expected reduction in annual volatility from 0.2451 to 0.2296. Further, the CAGR increases from 8.85% to 9.18%, leading to a higher Sharpe ratio of 0.4984.

Rather than experiencing signal decay when moving down in the prediction ranking, the RoBERTa-Sentiment model appears stable across the top tier of sectors. Conversely, the Non-Sentiment baseline model exhibits a decline in performance under the Top-2 setting, with CAGR decreasing by 0.36 percentage points. This suggests that purely price-based features lack the predictive breadth provided by narrative sentiment signals.

Table 5.8 – Results: Top-2 Long Strategy

Sector	CAGR	CAGR vs. Sector	Annual Volatility	Sharpe Ratio	IC Mean	IC t-stat	Avg. Turnover
RoBERTa Sentiment							
Top-2 Long	0.0918		0.2296	0.4984	0.0777	2.9906	0.3589
Equity REITs	0.0487	0.0431	0.2278	0.3235			
Industrial	0.1139	-0.0221	0.2455	0.5619			
Office	0.0009	0.0909	0.2719	0.1393			
Residential	0.0596	0.0322	0.2341	0.3667			
Retail	0.0271	0.0647	0.2932	0.2393			
GPT Sentiment							
Top-2 Long	0.0538		0.2358	0.3417	0.0054	0.2171	0.4134
Equity REITs	0.0487	0.0051	0.2278	0.3235			
Industrial	0.1139	-0.0601	0.2455	0.5619			
Office	0.0009	0.0529	0.2719	0.1393			
Residential	0.0596	-0.0058	0.2341	0.3667			
Retail	0.0271	0.0267	0.2932	0.2393			
Non-Sentiment							
Top-2 Long	0.0684		0.2360	0.4006	0.0636	2.4723	0.3411
Equity REITs	0.0487	0.0197	0.2278	0.3235			
Industrial	0.1139	-0.0455	0.2455	0.5619			
Office	0.0009	0.0675	0.2719	0.1393			
Residential	0.0596	0.0088	0.2341	0.3667			
Retail	0.0271	0.0413	0.2932	0.2393			

Note: This table presents the results of the Top-2 Long Strategy. The metrics presented for model evaluation are CAGR, Annual Volatility, Sharpe Ratio, and the IC Mean. The CAGR vs. Sector presents the relative outperformance of the proposed strategy versus the benchmark and the individual sectors. The IC t-stat depicts the statistical significance of the IC Mean and the Avg. Turnover resembles the average frequency of capital reallocation per week.

We also observe a substantial reduction in overall turnover, lowering the sensitivity of the proposed strategy to transaction costs. For the best-performing model, the break-even transaction costs for the Top-2 Long Strategy are approximately 21.6 bps.

Additionally, comparing the concentrated Top-1 with the diversified Top-2 strategy yields insights into the stability of predictive signals. While the non-sentiment baseline declines in performance when expanding the allocation, the RoBERTa-Sentiment model exhibits lower annual volatility without compromising – and even slightly improving – absolute returns. This observation suggests that the RoBERTa-Sentiment model captures a broader regime of outperforming sectors rather than only reacting to isolated weekly outliers.

The findings from the two trading strategies validate H2, demonstrating that sophisticated news-based sentiment indicators contain unique information relevant to REIT returns that translates into persistent economic gains.

5.7 Conclusion

This study provides comprehensive evidence on the short-term (weekly) predictability of U.S. REIT subsectors, with a particular focus on the role of investor sentiment in forecasting performance. We implement a strict OOS forecasting framework using linear (LASSO) and nonlinear (XGBoost) models across multiple specifications, comparing sentiment-enhanced models based on

measures derived from a fine-tuned RoBERTa model and a GPT-based model, relative to a non-sentiment benchmark.

While absolute return prediction remains highly elusive, relative sector outperformance exhibits economically meaningful predictability, particularly for the industrial and office sectors. Sentiment emerges as a key driver of model performance, indicating that news-based signals capture relative mispricing across subsectors, consistent with limits to arbitrage and heterogeneous beliefs. This cross-sectional mispricing gives rise to predictable return differentials that can be exploited through sector rotation strategies. Accordingly, the results suggest that sentiment should not be interpreted as a predictor of aggregate market movements, but rather as a signal for cross-sectional sector allocation. This finding is particularly relevant in the REIT context, where informational opacity and institutional ownership amplify the role of sentiment in the formation of asset prices.

Further, we exhibit that LLMs enable time- and cost-efficient processing of large amounts of textual data, making them a valuable tool for sentiment analysis. However, while measures of aggregate market sentiment derived from RoBERTa provide robust improvements over the non-sentiment benchmark, this does not hold for the GPT-based sector-specific sentiment, highlighting the importance of task-specific fine-tuning despite the general capabilities of generative LLMs.

While our analysis focuses on the REIT market, the underlying mechanisms are likely to extend to other equity markets. Thus, the findings are of interest to investors engaged in cross-sectional active capital reallocation strategies in the overall stock market. Future research may explore prompt optimization and domain adaptation of LLMs to improve the extraction of sector-specific sentiment signals and further examine their role in asset price formation.

5.8 Appendix

5.8.1 Appendix I – SentFiN 1.0 Prompt

You are a financial sentiment classifier for news headlines.

Task:

Given a financial news headline, classify the sentiment with respect to financial markets or the mentioned asset.

Use exactly one of these labels:

- positive

- negative

- neutral

Guidelines:

- Focus on the implied impact on prices, performance, or outlook.

- If the sentiment is unclear, mixed, or weak, use "neutral."

- Do not use any other labels.

Output format:

Respond with exactly one word, and nothing else:

- positive

- negative

- neutral

5.8.2 Appendix II – Sentiment Classification Prompt

You are a financial analyst specializing in U.S. REIT markets.

Analyze the {news_source} news article only based on its headline and {accompanying_text}.

Assign a sentiment score (-1.0 to +1.0 in 0.1 increments, or "n.r." if not relevant) for each FTSE EPRA Nareit U.S. Total Return REIT index:

- Equity REIT Index
- Office REIT Index
- Residential REIT Index
- Industrial REIT Index
- Retail REIT Index

Guidelines:

- Base sentiment only on the headline and the {accompanying_text}.
- Assign "n.r." if an index is not affected.
- Use positive values for positive impact (bullish/buy signal).
- Use negative values for negative impact (bearish/sell signal).
- Larger absolute values represent stronger expected impact.
- Weigh positive and negative impacts; assign 0.0 if neutral.
- Consider both direct (sector-specific) and indirect effects on each sector.
- Score each REIT index independently.

5.8.3 Appendix III – Mitigation of GPT's Look-Ahead Bias

The literature proposes several approaches to mitigate the memorization problem associated with LLMs. One strategy is to restrict the application to data published after the model's training cutoff (He et al., 2025; Lopez-Lira & Tang, 2023), while an alternative approach masks entity-specific information to prevent the model from recognizing the original article (Engelberg et al., 2025; Glasserman & Lin, 2024). Moreover, Lopez-Lira et al. (2025) and Didisheim et al. (2025) similarly assess look-ahead bias by analyzing the GPT models' capability to correctly recall the data of interest, such as index returns or article publication dates.

Since our analysis focuses on sector-specific sentiment extraction, where company names play a major role in determining sector relevance, we do not apply entity neutering. Instead, we follow Lopez-Lira et al.'s (2025) approach and prompt gpt-4.1-nano to estimate the publication date of each article in our dataset (see Appendix IV). However, it is important to note that we mask all date-related expressions before the analysis. This ensures that the task measures genuine memorization rather than straightforward date extraction from the text. Additionally, this prevents the model from drawing conclusions about the respective REIT returns based on the date in the final sentiment analysis of the articles. The results of this analysis are reported in Table 5.9.

The findings indicate that the GPT model exhibits a substantial deviation from the true publication date, with an average error of approximately 3,000 days across the three news sources. Although the proportions of correctly recalled years, months, or weeks are relatively low, they remain marginally above the accuracy expected under random guessing. To eliminate any residual risk of look-ahead bias in our dataset, we exclude all articles for which the model correctly recalled the publication week, as our objective is to forecast weekly returns based on extracted news sentiment. Moreover, we tested whether the model could recall the values of any of the five REIT indices analyzed in this study. We find no evidence that gpt-4.1-nano memorized any of the REIT index series. Instead, the results displayed in Table 5.10 indicate random guessing. Additionally, we test whether the model memorizes the original monthly Nareit return series, considering both price and total return, and obtain similar results.

Table 5.9 – Evaluation of Date Recall Accuracy for Historical News Articles

	Mean Days Diff.	Mean Abs. Days Diff.	Year Acc.	Month and Year Acc.	Week and Year Acc.	Exact Date Acc.	Confidence Calibration	Num. Obs.
S&P MI REIT News								
Pre-cutoff	2,739.96	2,743.68	4.13%	1.12%	0.44%	0.08%	0.06	49,005
Post-cutoff	-482.43	482.43	0.17%	0.00%	0.00%	0.00%	-0.07	1,172
S&P MI Macroeconomic & Policy News								
Pre-cutoff	2,527.75	2,537.90	8.28%	2.24%	1.13%	0.44%	0.15	132,493
Post-cutoff	-508.65	508.99	1.26%	0.41%	0.21%	0.04%	-0.05	4,832
CoStar News								
Pre-cutoff	2,289.31	2,301.48	7.55%	1.10%	0.33%	0.07%	0.07	121,801
Post-cutoff	-486.44	486.57	0.49%	0.07%	0.03%	0.01%	-0.06	11,916

Note: Following Lopez-Lira et al. (2025), this table reports gpt-4.1-nano's ability to recall publication dates of historical articles based on their headline and subheading/full article text from October 24, 2009, to October 31, 2025. The model was prompted to state the estimated date alongside a confidence score (0–100). All metrics are reported separately for periods before and after the model's knowledge cutoff date (June 1, 2024). Mean Days Difference is the average signed prediction error in days. Mean Absolute Days Difference is the absolute counterpart. Year Accuracy, Month and Year Accuracy, Week and Year Accuracy, and Exact Date Accuracy report the share of observations where the LLM correctly recalled the respective time unit. Confidence Calibration is the Pearson correlation between the model's confidence score and the negative absolute days difference, where positive values indicate alignment between confidence and accuracy. First instances of valid date format responses occurred across all datasets, all within the pre-cutoff date.

Table 5.10 – Evaluation of Recall Accuracy for FTSE EPRA Nareit REIT Indices

	MPE (%)	MAPE (%)	Directional Accuracy (%)	Confidence Calibration	Num. Obs.
Daily Levels: Pre-cutoff					
Equity REITs Index	-67.39	71.88	46.35	0.02	3,978
Office REITs Index	-40.35	102.47	47.62	-0.70	3,978
Residential REITs Index	-75.09	95.77	46.12	-0.63	3,978
Industrial REITs Index	5.34	88.63	48.02	-0.65	3,978
Retail REITs Index	-59.68	104.42	46.21	-0.62	3,978
Daily Levels: Post-cutoff					
Equity REITs Index	-49.37	56.73	48.17	0.22	369
Office REITs Index	485.96	490.15	56.06	-0.52	369
Residential REITs Index	-29.67	68.13	51.55	-0.19	369
Industrial REITs Index	188.11	202.66	51.27	0.24	369
Retail REITs Index	63.91	76.00	47.89	0.02	369
Weekly Returns: Pre-cutoff					
Equity REITs Index	-72.22	183.84	74.88	-0.03	805
Office REITs Index	-100.32	125.54	75.00	-0.06	805
Residential REITs Index	-195.90	276.39	74.50	-0.02	805
Industrial REITs Index	174.92	426.00	74.75	0.02	805
Retail REITs Index	-105.57	156.56	75.50	-0.05	805
Weekly Returns: Post-cutoff					
Equity REITs Index	14.88	282.60	76.71	0.12	74
Office REITs Index	-103.80	103.80	75.34	0.11	74
Residential REITs Index	-96.07	100.76	75.34	0.03	74
Industrial REITs Index	-171.04	204.80	79.45	0.18	74
Retail REITs Index	-58.38	133.56	83.56	0.19	74
Weekly Directional Changes: Pre-cutoff					
Equity REITs Index			58.14	0.01	805
Office REITs Index			56.27	0.01	805
Residential REITs Index			57.39	-0.01	805
Industrial REITs Index			57.39	0.01	805
Retail REITs Index			55.65	-0.02	805
Weekly Directional Changes: Post-cutoff					
Equity REITs Index			59.46	-0.17	74
Office REITs Index			58.11	-0.10	74
Residential REITs Index			52.70	-0.12	74
Industrial REITs Index			56.76	-0.19	74
Retail REITs Index			59.46	-0.05	74

Note: Following the methodology of Lopez-Lira et al. (2025), this table reports relevant evaluation metrics assessing gpt-4.1-nano's ability to recall index and return values of five FTSE EPRA Nareit REIT indices from January 1, 2009, to October 31, 2025. The model was prompted to state the respective value together with its confidence level (on a scale from 0 to 100) regarding its answer. All metrics are reported for the periods before and after the model's knowledge cutoff date (June 1, 2024). For Daily Levels, the model was asked to state the index value for each trading day, while it was asked to state the percentage change of a week (i.e., the end-of-week return) for Weekly Returns. For Weekly Directional Changes, the model was prompted to answer whether the index was up or down at the end of the respective week. Mean percentage error (MPE) is computed as the mean difference between predicted and actual return divided by the actual return, while mean absolute percentage error (MAPE) uses the absolute percentage errors. Confidence Calibration is computed as the correlation between the model's reported confidence score and the absolute percentage error. The Directional Accuracy is defined as the share of correctly classified up/down movements. For Daily Levels, it is calculated by comparing the true direction with the predicted direction, defined as the sign of the change between the model's predicted index level and the previous day's true index level.

5.8.4 Appendix IV – Recall Accuracy Prompts

All prompts follow the exact same design as outlined in Lopez-Lira et al. (2025).

News articles

Here is an article from {news_source}:

Title: {headline}

{accompanying_text_item}: {accompanying_text}

What is the date of this article? Provide your best estimate.

Provide a precise date. Indicate your level of confidence. Format as a JSON object with the following fields:

• answer: The precise date in the format \"mm/dd/yyyy\".

• confidence: A number between 0 and 100 indicating your confidence in the answer.

Daily levels

What was the {index_name} closing value on {date_str}? Provide your best estimate.

Provide a precise numerical answer. Indicate your level of confidence. Format as a JSON object with the following fields:

• answer: The precise numerical answer to the question. No strings.

• confidence: A number between 0 and 100 indicating your confidence in the answer.

Weekly returns

By what percentage did the {index_name} change for the week ending {date_str}? Provide your best estimate.

Provide a precise numerical answer. Indicate your level of confidence. Format as a JSON object with the following fields:

• answer: The precise numerical answer to the question. No strings.

• confidence: A number between 0 and 100 indicating your confidence in the answer.

Weekly directional changes

Was the {index_name} up or down for the week ending {date_str}? Provide your best estimate.

Provide an answer that is either \"up\" or \"down\". Indicate your level of confidence. Format as a JSON object with the following fields:

• answer: An answer to the question that is either \"up\" or \"down\".

• confidence: A number between 0 and 100 indicating your confidence in the answer.

5.9 References

Amin, M. M., Cambria, E., & Schuller, B. W. (2023). Will affective computing emerge from foundation models and general artificial intelligence? A first evaluation of ChatGPT. *IEEE Intelligent Systems*, *38*(2), 15–23. https://doi.org/10.1109/MIS.2023.3254179

Antweiler, W., & Frank, M. Z. (2004). Is all that talk just noise? The information content of internet stock message boards. *The Journal of Finance*, *59*(3), 1259–1294. https://doi.org/10.1111/j.1540-6261.2004.00662.x

Araci, D. (2019). *FinBERT: Financial sentiment analysis with pre-trained language models*. arXiv. https://doi.org/10.48550/arXiv.1908.10063

Baker, M., & Wurgler, J. (2006). Investor sentiment and the cross-section of stock returns. *The Journal of Finance*, *61*(4), 1645–1680. https://doi.org/10.1111/j.1540-6261.2006.00885.x

Baker, M., & Wurgler, J. (2007). Investor sentiment in the stock market. *Journal of Economic Perspectives*, *21*(2), 129–151. https://doi.org/10.1257/jep.21.2.129

Bond, S., Guo, H., & Yang, C. (2025). Systematic mispricing: Evidence from real estate markets. *The Journal of Real Estate Finance and Economics*, *71*(3), 420–452. https://doi.org/10.1007/s11146-021-09883-9

Boyd, J. H., Hu, J., & Jagannathan, R. (2005). The stock market's reaction to unemployment news: Why bad news is usually good for stocks. *The Journal of Finance*, *60*(2), 649–672. https://doi.org/10.1111/j.1540-6261.2005.00742.x

Brooks, C., & Tsolacos, S. (2003). International evidence on the predictability of returns to securitized real estate assets: Econometric models versus neural networks. *Journal of Property Research*, *20*(2), 133–155. https://doi.org/10.1080/0959991032000109517

Brown, G. W., & Cliff, M. T. (2004). Investor sentiment and the near-term stock market. *Journal of Empirical Finance*, *11*(1), 1–27. https://doi.org/10.1016/j.jempfin.2002.12.001

Brown, G. W., & Cliff, M. T. (2005). Investor sentiment and asset valuation. *The Journal of Business*, *78*(2), 405–440. https://doi.org/10.1086/427633

Brown, T. B., Mann, B., Ryder, N., Subbiah, M., Kaplan, J., Dhariwal, P., Neelakantan, A., Shyam, P., Sastry, G., Askell, A., Agarwal, S., Herbert-Voss, A., Krueger, G., Henighan, T., Child, R., Ramesh, A., Ziegler, D. M., Wu, J. [Jeffrey], Winter, C., . . . Amodei, D. (2020). Language models are few-shot learners. In H. Larochelle, M. Ranzato, R. Hadsell, M. F. Balcan, & H. Lin (Eds.), *Advances in Neural Information Processing Systems* (Vol. 33, pp. 1877–1901). Curran Associates.

https://papers.nips.cc/paper_files/paper/2020/file/1457c0d6bfcb4967418bfb8ac142f64a-Paper.pdf

Cabrera, G., Díaz, J. D., & Hansen, E. (2026). Real estate returns and the macroeconomy: Insights from big data in the US, Canada, and the UK. *The Journal of Real Estate Finance and Economics*. Advance online publication. https://doi.org/10.1007/s11146-025-10045-4

Campbell, J. Y., & Thompson, S. B. (2008). Predicting excess stock returns out of sample: Can anything beat the historical average? *The Review of Financial Studies*, *21*(4), 1509–1531. https://doi.org/10.1093/rfs/hhm055

Carstens, R., & Freybote, J. (2021). Can the textual tone in REIT financial statements improve the information environment for commercial real estate investors? An investigation. *Journal of Real Estate Research*, *43*(3), 335–354. https://doi.org/10.1080/08965803.2021.1985923

Chen, T., & Guestrin, C. (2016). XGBoost: A scalable tree boosting system. In *Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining* (pp. 785–794). Association for Computing Machinery. https://doi.org/10.1145/2939672.2939785

Chiang, S. L., & Tsai, M. S. (2023). Analyses for the effects of investor sentiment on the price adjustment behaviors for stock market and REIT market. *International Review of Economics & Finance*, *86*, 425–439. https://doi.org/10.1016/j.iref.2023.03.007

Chiu, I.-C., & Hung, M.-W. (2025). Finance-specific large language models: Advancing sentiment analysis and return prediction with LLaMA 2. *Pacific-Basin Finance Journal*, *90*, Article 102632. https://doi.org/10.1016/j.pacfin.2024.102632

Churchill, A., Pichika, S., Xu, C., & Liu, Y. [Ying]. (2026). GPT models for text annotation: An empirical exploration in public policy research. *Policy Studies Journal*, *54*(1), Article e70034. https://doi.org/10.1111/psj.70034

Da, Z., Engelberg, J., & Gao, P. (2015). The sum of all FEARS investor sentiment and asset prices. *The Review of Financial Studies*, *28*(1), 1–32. https://doi.org/10.1093/rfs/hhu072

Dangl, T., & Halling, M. (2012). Predictive regressions with time-varying coefficients. *Journal of Financial Economics*, *106*(1), 157–181. https://doi.org/10.1016/j.jfineco.2012.04.003

Das, P. K., Freybote, J., & Marcato, G. (2015). An investigation into sentiment-induced institutional trading behavior and asset pricing in the REIT market. *The Journal of Real Estate Finance and Economics*, *51*(2), 160–189. https://doi.org/10.1007/s11146-014-9490-z

De Long, J. B., Shleifer, A., Summers, L. H., & Waldmann, R. J. (1990). Noise trader risk in financial markets. *Journal of Political Economy*, *98*(4), 703–738. https://doi.org/10.1086/261703

Demiralay, S., & Kilincarslan, E. (2024). Uncertainty measures and sector-specific REITs in a regime-switching environment. *The Journal of Real Estate Finance and Economics*, *69*, 545–584. https://doi.org/10.1007/s11146-022-09898-w

Devlin, J., Chang, M.-W., Lee, K., & Toutanova, K. (2019). BERT: Pre-training of deep bidirectional transformers for language understanding. In J. Burstein, C. Doran, & T. Solorio (Eds.), *Proceedings of the 2019 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies, Volume 1 (Long and Short Papers)* (pp. 4171–4186). Association for Computational Linguistics. https://doi.org/10.18653/v1/N19-1423

Didisheim, A., Fraschini, M., & Somoza, L. (2025). AI's predictable memory in financial analysis. *Economics Letters*, *256*, Article 112602. https://doi.org/10.1016/j.econlet.2025.112602

Donaldson, R. G., & Kamstra, M. (1996). Forecast combining with neural networks. *Journal of Forecasting*, *15*, 49–61. https://doi.org/10.1002/(SICI)1099-131X(199601)15:1<49::AID-FOR604>3.0.CO;2-2

Engelberg, J., Manela, A., Mullins, W., & Vulicevic, L. (2025). *Entity neutering* [Working paper]. SSRN. https://doi.org/10.2139/ssrn.5182756

Fama, E. F., & French, K. R. (1992). The cross-section of expected stock returns. *The Journal of Finance*, *47*(2), 427–465. https://doi.org/10.2307/2329112

Fatouros, G., Soldatos, J., Kouroumali, K., Makridis, G., & Kyriazis, D. (2023). Transforming sentiment analysis in the financial domain with ChatGPT. *Machine Learning with Applications*, *14*, Article 100508. https://doi.org/10.1016/j.mlwa.2023.100508

Feldman, R., Govindaraj, S., Livnat, J., & Segal, B. (2010). Management's tone change, post earnings announcement drift and accruals. *Review of Accounting Studies*, *15*(4), 915–953. https://doi.org/10.1007/s11142-009-9111-x

Ferreira, M. A., & Santa-Clara, P. (2011). Forecasting stock market returns: The sum of the parts is more than the whole. *Journal of Financial Economics*, *100*(3), 514–537. https://doi.org/10.1016/j.jfineco.2011.02.003

Frankel, R., Jennings, J., & Lee, J. [Joshua]. (2022). Disclosure sentiment: Machine learning vs. dictionary methods. *Management Science*, *68*(7), 5514–5532. https://doi.org/10.1287/mnsc.2021.4156

Friedman, J. H. (2001). Greedy function approximation: A gradient boosting machine. *The Annals of Statistics*, *29*(5), 1189–1232. https://doi.org/10.1214/aos/1013203451

García, D. (2013). Sentiment during recessions. *The Journal of Finance*, *68*(3), 1267–1300. https://doi.org/10.1111/jofi.12027

Gilardi, F., Alizadeh, M., & Kubli, M. (2023). ChatGPT outperforms crowd workers for text-annotation tasks. *Proceedings of the National Academy of Sciences of the United States of America*, *120*(30), Article e2305016120. https://doi.org/10.1073/pnas.2305016120

Gilbert, T. (2011). Information aggregation around macroeconomic announcements: Revisions matter. *Journal of Financial Economics*, 101(1), 114–131. https://doi.org/10.1016/j.jfineco.2011.02.013

Glasserman, P., & Lin, C. (2024). Assessing look-ahead bias in stock return predictions generated by GPT sentiment analysis. *The Journal of Financial Data Science*, *6*(1), 25–42. https://doi.org/10.3905/jfds.2023.1.143

Gu, C., & Kurov, A. (2020). Informational role of social media: Evidence from Twitter sentiment. *Journal of Banking & Finance*, *121*, Article 105969. https://doi.org/10.1016/j.jbankfin.2020.105969

Gu, S., Kelly, B., & Xiu, D. (2020). Empirical asset pricing via machine learning. *The Review of Financial Studies*, *33*(5), 2223–2273. https://doi.org/10.1093/rfs/hhaa009

Guevarra, J., & Hudgins, C. (2024). *Foreign institutional investor ownership in US REITs grows*. S&P Global Market Intelligence. https://www.spglobal.com/market-intelligence/en/news-insights/articles/2024/3/foreign-institutional-investor-ownership-in-us-reits-grows-80637545

Hastie, T., Tibshirani, R., & Friedman, J. (2009). *The elements of statistical learning: Data mining, inference, and prediction* (2nd ed.). Springer. https://doi.org/10.1007/978-0-387-84858-7

He, S., Lv, L., Manela, A., & Wu, J. [Jimmy]. (2025). *Chronologically consistent large language models*. arXiv. https://doi.org/10.48550/arXiv.2502.21206

Hirshleifer, D., & Sheng, J. (2022). Macro news and micro news: Complements or substitutes? *Journal of Financial Economics*, *145*(3), 1006–1024. https://doi.org/10.1016/j.jfineco.2021.09.012

Ho, K.-Y., Shi, Y., & Zhang, Z. (2013). How does news sentiment impact asset volatility? Evidence from long memory and regime-switching approaches. *The North American Journal of Economics and Finance*, *26*, 436–456. https://doi.org/10.1016/j.najef.2013.02.015

Huerta, D., Jackson, D. O., & Ngo, T. (2015). Categorizing sentiment and its impact on REIT returns. *Managerial Finance*, *41*(9), 958–973. https://doi.org/10.1108/MF-06-2014-0164

Huerta-Sanchez, D., & Escobari, D. (2018). Changes in sentiment on REIT industry excess returns and volatility. *Financial Markets and Portfolio Management*, *32*(3), 239–274. https://doi.org/10.1007/s11408-018-0312-9

Jenett, H., Nagl, C., Nagl, M., Price, S. M., & Schaefers, W. (2026). Dynamics of REIT returns and volatility: Analyzing time-varying drivers through an explainable machine learning approach. *The Journal of Real Estate Finance and Economics*, *72*(1), 1–40. https://doi.org/10.1007/s11146-025-10016-9

Kempf, E., Manconi, A., & Spalt, O. (2017). Distracted shareholders and corporate actions. *The Review of Financial Studies*, *30*(5), 1660–1695. https://doi.org/10.1093/rfs/hhw082

Kim, J. H. (2023). What if ChatGPT were a quant asset manager. *Finance Research Letters*, *58*, Article 104580. https://doi.org/10.1016/j.frl.2023.104580

Kirtac, K., & Germano, G. (2024). Sentiment trading with large language models. *Finance Research Letters*, *62*, Article 105227. https://doi.org/10.1016/j.frl.2024.105227

Ko, H., & Lee, J. [Jaewook]. (2024). Can ChatGPT improve investment decisions? *From a portfolio management perspective. Finance Research Letters*, *64*, Article 105433. https://doi.org/10.1016/j.frl.2024.105433

Lee, C. M. C., Shleifer, A., & Thaler, R. H. (1991). Investor sentiment and the closed-end fund puzzle. *The Journal of Finance*, *46*(1), 75–109. https://doi.org/10.1111/j.1540-6261.1991.tb03746.x

Lee, W. Y., Jiang, C. X., & Indro, D. C. (2002). Stock market volatility, excess returns, and the role of investor sentiment. *Journal of Banking & Finance*, *26*(12), 2277–2299. https://doi.org/10.1016/S0378-4266(01)00202-3

Lemmon, M., & Portniaguina, E. (2006). Consumer confidence and asset prices: Some empirical evidence. *The Review of Financial Studies*, *19*(4), 1499–1529. https://doi.org/10.1093/rfs/hhj038

Leow, K., & Lindenthal, T. (2025). Enhancing real estate investment trust return forecasts using machine learning. *Real Estate Economics*, *53*(3), 574–606. https://doi.org/10.1111/1540-6229.12527

Ling, D. C., Naranjo, A., & Ryngaert, M. D. (2000). The predictability of equity REIT returns: Time variation and economic significance. *The Journal of Real Estate Finance and Economics*, *20*(2), 117–136. https://doi.org/10.1023/a:1007821103728

Liu, T., & Shi, Y. (2025). News sentiment and investment risk management: Innovative evidence from the large language models. *Economics Letters*, *247*, Article 112124. https://doi.org/10.1016/j.econlet.2024.112124

Liu, Y. [Yinhan], Ott, M., Goyal, N., Du, J., Joshi, M., Chen, D., Levy, O., Lewis, M., Zettlemoyer, L., & Stoyanov, V. (2019). *RoBERTa: A robustly optimized BERT pretraining approach*. arXiv. https://doi.org/10.48550/arxiv.1907.11692

Lopez-Lira, A., & Tang, Y. (2023). *Can ChatGPT forecast stock price movements? Return predictability and large language models* [Working paper]. SSRN. https://doi.org/10.2139/ssrn.4412788

Lopez-Lira, A., Tang, Y., & Zhu, M. (2025). *The memorization problem: Can we trust LLMs' economic forecasts?* arXiv. https://doi.org/10.48550/arXiv.2504.14765

Loughran, T., & McDonald, B. (2011). When is a liability not a liability? Textual analysis, dictionaries, and 10-Ks. *The Journal of Finance*, *66*(1), 35–65. https://doi.org/10.1111/j.1540-6261.2010.01625.x

Ma, F., Lyu, Z., & Li, H. (2024). Can ChatGPT predict Chinese equity premiums? *Finance Research Letters*, *65*, Article 105631. https://doi.org/10.1016/j.frl.2024.105631

Mahmoudi, N., Docherty, P., & Melia, A. (2022). Firm-level investor sentiment and corporate announcement returns. *Journal of Banking & Finance*, *144*, Article 106586. https://doi.org/10.1016/j.jbankfin.2022.106586

Malo, P., Sinha, A., Korhonen, P., Wallenius, J., & Takala, P. (2014). Good debt or bad debt: Detecting semantic orientations in economic texts. *Journal of the Association for Information Science and Technology*, *65*(4), 782–796. https://doi.org/10.1002/asi.23062

Marcato, G., & Nanda, A. (2016). Information content and forecasting ability of sentiment indicators: Case of real estate market. *Journal of Real Estate Research*, *38*(2), 165–204. https://doi.org/10.1080/10835547.2016.12091442

Mei, J., & Gao, B. (1995). Price reversal, transaction costs, and arbitrage profits in the real estate securities market. *The Journal of Real Estate Finance and Economics*, *11*(2), 153–165. https://doi.org/10.1007/BF01098659

Neal, R., & Wheatley, S. M. (1998). Do measures of investor sentiment predict returns? *The Journal of Financial and Quantitative Analysis*, *33*(4), 523–547. https://doi.org/10.2307/2331130

Oehler, A., & Horn, M. (2024). Does ChatGPT provide better advice than robo-advisors? *Finance Research Letters*, *60*, Article 104898. https://doi.org/10.1016/j.frl.2023.104898

Patelis, A. D. (1997). Stock return predictability and the role of monetary policy. *The Journal of Finance*, *52*(5), 1951–1972. https://doi.org/10.1111/j.1540-6261.1997.tb02747.x

Pelster, M., & Val, J. (2024). Can ChatGPT assist in picking stocks? *Finance Research Letters*, *59*, Article 104786. https://doi.org/10.1016/j.frl.2023.104786

Peng, L., & Xiong, W. (2006). Investor attention, overconfidence and category learning. *Journal of Financial Economics*, *80*(3), 563–602. https://doi.org/10.1016/j.jfineco.2005.05.003

Pesaran, M. H., & Timmermann, A. (1994). Forecasting stock returns an examination of stock market trading in the presence of transaction costs. *Journal of Forecasting*, *13*(4), 335–367. https://doi.org/10.1002/for.3980130402

Plößl, F., Paulus, N. M., & Just, T. (2023). Trade vs. daily press: The role of news coverage and sentiment in real estate market analysis. *Journal of Property Research*, *40*(4), 333–364. https://doi.org/10.1080/09599916.2023.2199764

Price, S. M., Doran, J. S., Peterson, D. R., & Bliss, B. A. (2012). Earnings conference calls and stock returns: The incremental informativeness of textual tone. *Journal of Banking & Finance*, *36*(4), 992–1011. https://doi.org/10.1016/j.jbankfin.2011.10.013

Qi, M., & Maddala, G. S. (1999). Economic factors and the stock market: A new perspective. *Journal of Forecasting*, *18*(3), 151–166. https://doi.org/10.1002/(SICI)1099-131X(199905)18:3<151::AID-FOR716>3.0.CO;2-V

Radford, A., Narasimhan, K., Salimans, T., & Sutskever, I. (2018). *Improving language understanding by generative pre-training*. OpenAI. https://cdn.openai.com/research-covers/language-unsupervised/language_understanding_paper.pdf

Rapach, D. E., Strauss, J. K., & Zhou, G. (2010). Out-of-sample equity premium prediction: Combination forecasts and links to the real economy. *The Review of Financial Studies*, *23*(2), 821–862. https://doi.org/10.1093/rfs/hhp063

Rathje, S., Mirea, D.-M., Sucholutsky, I., Marjieh, R., Robertson, C. E., & Van Bavel, J. J. (2024). GPT is an effective tool for multilingual psychological text analysis. *Proceedings of*

the National Academy of Sciences of the United States of America, *121*(34), Article e2308950121. https://doi.org/10.1073/pnas.2308950121

Renault, T. (2017). Intraday online investor sentiment and return patterns in the U.S. stock market. *Journal of Banking & Finance*, *84*, 25–40. https://doi.org/10.1016/j.jbankfin.2017.07.002

Rochdi, K., & Dietzel, M. (2015). Outperforming the benchmark: online information demand and REIT market performance. *Journal of Property Investment & Finance*, *33*(2), 169–195. https://doi.org/10.1108/JPIF-11-2014-0069

Ruscheinsky, J. R., Lang, M., & Schäfers, W. (2018). Real estate media sentiment through textual analysis. *Journal of Property Investment & Finance*, *36*(5), 410–428. https://doi.org/10.1108/JPIF-07-2017-0050

Schneider, C. J., & Yilmaz, Y. (2025). Stock portfolio selection based on risk appetite: Evidence from ChatGPT. *Finance Research Letters*, *82*, Article 107517. https://doi.org/10.1016/j.frl.2025.107517

Siano, F., & Wysocki, P. (2021). Transfer learning and textual analysis of accounting disclosures: Applying big data methods to small(er) datasets. *Accounting Horizons*, *35*(3), 217–244. https://doi.org/10.2308/HORIZONS-19-161

Sinha, A., Kedas, S., Kumar, R., & Malo, P. (2022). SEntFiN 1.0: Entity-aware sentiment analysis for financial news. *Journal of the Association for Information Science and Technology*, *73*(9), 1314–1335. https://doi.org/10.1002/asi.24634

Stambaugh, R. F., Yu, J., & Yuan, Y. (2012). The short of it: Investor sentiment and anomalies. *Journal of Financial Economics*, *104*(2), 288–302. https://doi.org/10.1016/j.jfineco.2011.12.001

Sun, L., Najand, M., & Shen, J. (2016). Stock return predictability and investor sentiment: A high-frequency perspective. *Journal of Banking & Finance*, *73*, 147–164. https://doi.org/10.1016/j.jbankfin.2016.09.010

Suzuki, M., Ishikawa, Y., Teraguchi, M., & Sakaji, H. (2025). Sentiment works in small-cap stocks: Japanese stock's sentiment with language models. *International Journal of Information Management Data Insights*, *5*(1), Article 100318. https://doi.org/10.1016/j.jjimei.2024.100318

Swaminathan, B. (1996). Time-varying expected small firm returns and closed-end fund discounts. *The Review of Financial Studies*, *9*(3), 845–887. https://doi.org/10.1093/rfs/9.3.845

Tetlock, P. C. (2007). Giving content to investor sentiment: The role of media in the stock market. *The Journal of Finance*, *62*(3), 1139–1168. https://doi.org/10.1111/j.1540-6261.2007.01232.x

Tetlock, P. C., Saar-Tsechansky, M., & Macskassy, S. (2008). More than words: Quantifying language to measure firms' fundamentals. *The Journal of Finance*, *63*(3), 1437–1467. https://doi.org/10.1111/j.1540-6261.2008.01362.x

Tibshirani, R. (1996). Regression shrinkage and selection via the lasso. *Journal of the Royal Statistical Society Series B: Statistical Methodology*, *58*(1), 267–288. https://doi.org/10.1111/j.2517-6161.1996.tb02080.x

Vaswani, A., Shazeer, N., Parmar, N., Uszkoreit, J., Jones, L., Gomez, A. N., Kaiser, L., & Polosukhin, I. (2017). *Attention is all you need*. arXiv. https://doi.org/10.48550/arXiv.1706.03762

Welch, I., & Goyal, A. (2008). A comprehensive look at the empirical performance of equity premium prediction. *The Review of Financial Studies*, *21*(4), 1455–1508. https://doi.org/10.1093/rfs/hhm014

Zhang, W., Gong, X., Wang, C., & Ye, X. (2021). Predicting stock market volatility based on textual sentiment: A nonlinear analysis. *Journal of Forecasting*, *40*(8), 1479–1500. https://doi.org/10.1002/for.2777

6 Conclusion

6.1 Executive Summary

The following section provides a concise summary of the four papers that constitute this cumulative dissertation. For each paper, the underlying research problem and objective are outlined, followed by the data and methodology employed, and the key results along with their contribution to science and practice.

Paper 1: Social Media and Real Estate: Do Twitter Users Predict REIT Performance?

Problems and Objective

Considering Shiller's (2017) narrative economics, information that investors read can influence their decision-making and, ultimately, market outcomes. The sentiment expressed in such information is therefore a relevant factor for understanding market dynamics, particularly in direct real estate markets, which are frequently deemed inefficient, illiquid, and characterized by significant information asymmetries (Gallimore & Gray, 2002). These particularities are expected to translate to the REIT market through its hybrid nature, combining equity market trading with exposure to the direct real estate market. Consistent with this expectation, the literature over the last decade has provided clear evidence for a dynamic relationship between text-based sentiment and REIT market outcomes, drawing on corporate disclosures (e.g., Carstens & Freybote, 2019; Koelbl, 2020) and news-based sources targeting real estate (e.g., Hausler et al., 2018; Ruscheinsky et al., 2018).

In contrast to these studies, which primarily rely on formal texts, social media has only been used occasionally in recent financial literature (e.g., Li et al., 2018; Sprenger et al., 2014), and even more rarely in the real estate context (e.g., Hannum et al., 2019; Zamani & Schwartz, 2017). This is despite the development of such platforms into continuously available, high-frequency channels of investor communication. The limited use of social media in academic research can be partly attributed to the inherent complexity of such unstructured data (Maynard et al., 2012). On social media, informal language, abbreviations, and slang terms complicate sentiment extraction, making careful preprocessing and adequate classifier choice essential.

The objective of this paper is to examine whether social media sentiment influences future U.S. REIT market returns. In doing so, it develops a framework for data selection and preprocessing and systematically compares multiple sentiment classifiers regarding their classification performance and predictive power.

Methodology and Data

The textual dataset initially contains 4.97 million Twitter messages from 2013 to 2022. These messages are collected using ticker symbols and other firm identifiers of publicly traded U.S. REITs. We apply an extensive filtering and data-cleaning approach to overcome the hurdles introduced by social media data, including classical (e.g., lowercasing) and advanced preprocessing (e.g., abbreviation expansion and slang conversion). The data is then filtered for semantic relevance using a Global Vectors (GloVe) model and further refined through interaction-based filtering, resulting in a final dataset of approximately 210,000 messages. For sentiment classification, we compare three approaches: a dictionary-based method, a support vector machine (SVM) model, and a long short-term memory (LSTM)-based neural network.

The FTSE Nareit All REITs Total Return Index serves as a proxy for the U.S. REIT market. To examine the sentiment-return relationship, we use a vector autoregression (VAR) model that includes total returns and the derived sentiment as endogenous variables, allowing for the assessment of bidirectional relationships and varying lag structures. In addition, macroeconomic control variables are incorporated to improve the accuracy and robustness of the model.

Results and Their Contribution to Science and Practice

The findings reveal a two-phase structure that maps onto the hybrid nature of REITs, with short-term reactions tied to stock market movements and long-term reactions of up to six to seven months acting as early indicators of changes in the underlying real estate market. This relationship is found to be unidirectional, with sentiment leading returns but not vice versa. For investors, this distinction matters, as the short-term component reflects reversal patterns driven by speculative dynamics, while the longer-term component carries information that may be relevant for positioning decisions over multi-month horizons. The comparison of sentiment classifiers points to a trade-off between accessibility and depth. The dictionary-based approach captures the same directional lag structure as more sophisticated models and may be sufficient for an initial assessment. However, the LSTM-derived sentiment exhibits the strongest and most significant relationship with REIT returns, making it more appropriate for contexts where signal precision matters. All three classifiers consistently identify the same lag structure, reinforcing the robustness of the underlying sentiment-return relationship rather than attributing it to a specific extraction method.

This paper expands the scope of textual sentiment analysis in real estate research by introducing social media as a continuously available, real-time alternative to formal text sources, and is the first to systematically examine its relationship with REIT market returns. Beyond the REIT-Twitter context, the framework for data selection, preprocessing, and sentiment classification of informal

text is transferable to other asset classes and informal text sources, offering a replicable methodological foundation for future sentiment research in financial markets.

Paper 2: Fundamentals or Noise? The Informative Value of Twitter Sentiment for REIT Returns

Problems and Objective

Recent literature has documented that social media sentiment contains predictive information for stock market returns (e.g., Sprenger et al., 2014), and initial evidence suggests that this also extends to the U.S. REIT market (Paulus et al., 2025).[1] However, whether this explanatory power reflects fundamental information or noise remains an open question. This distinction is particularly relevant in the context of social media, where investor communication can reflect both valuable information and noisier behavioral components. For REITs, the dynamic is especially pronounced due to their limited retail visibility on social media, given that Twitter users generally gravitate toward large, well-known, and attention-grabbing stocks (Barber & Odean, 2008; Cookson et al., 2024; Welch, 2022). That social media sentiment is nonetheless informative for REIT returns is, therefore, not self-evident, raising the question of what drives this informativeness. This question becomes even more relevant during the COVID-19 period, when attention shifted more strongly toward the real estate industry due to increased media coverage (Ozik et al., 2021), potentially altering the composition of active users on the platform.

This paper aims to decompose social media sentiment into a fundamentals-based and a non-fundamental component and to examine whether the component driving the sentiment-return relationship differs in the pre-COVID and COVID periods.

Methodology and Data

This study uses REIT-related Twitter data and covers the period from 2014 to 2022, allowing for a comparison between the pre-COVID and COVID periods. We define the pre-COVID period from January 2014 to January 2020 and the COVID period from February 2020 to the end of 2022. The methodological approach follows the data preparation framework established in Paulus et al. (2025), while data filtering and relevance selection are further refined. Initially, our dataset contained about 3.9 million messages, which were reduced to approximately 175,000 messages after filtering. For sentiment classification, we use a fine-tuned Robustly Optimized BERT Pretraining Approach (RoBERTa) model to distinguish between positive, negative, and neutral messages.

After index construction, we isolate the fundamentals-based component of sentiment by regressing the sentiment variable on macroeconomic, capital market, and commercial real estate market factors (Ling et al., 2014). This decomposition is estimated separately for the pre-COVID

[1] Paulus et al. (2025) corresponds to the published version of Paper 1.

and COVID periods, yielding a fundamentals-based component and a non-fundamental residual component for each period.

Finally, we use a vector autoregression (VAR) model to investigate how the two derived sentiment components explain U.S. REIT returns, proxied by the FTSE Nareit All Equity REITs Total Return Index, in the pre-COVID and COVID periods. In addition, we include macroeconomic control variables to improve model robustness.

Results and Their Contribution to Science and Practice

The decomposition reveals that the sentiment-return relationship is most pronounced for pessimistic sentiment, with its non-fundamental component exhibiting explanatory power in the pre-COVID period, whereas its fundamentals-based component becomes the dominant factor during the COVID period. The COVID period serves not only as a distinct economic regime but also as evidence that the informational content of social media sentiment is not constant and appears to change with the composition of active users on the platform. These findings add to the existing REIT sentiment literature by providing initial evidence on the fundamental drivers of REIT-specific social media sentiment and by demonstrating that aggregated sentiment measures can mask a regime-dependent shift between fundamentals-based and non-fundamental components.

For investors, the weight assigned to social media sentiment in allocation decisions should therefore reflect the prevailing market environment. In periods of heightened attention and market stress, social media sentiment tracks fundamentals more closely and can serve as a timely indicator. In calmer market phases, the signal is more strongly driven by non-fundamental factors and should be interpreted with greater caution. For researchers, the results imply that treating sentiment as a static variable in return regressions spanning a long horizon may obscure economically meaningful variations in its informational content.

Paper 3: Signals or Stories? ESG Sentiment and Investor Reactions in REIT Corporate Disclosure

Problems and Objective

ESG considerations have become increasingly relevant for corporate strategy and investment decisions in listed real estate markets. REITs operate at the intersection of sustainability challenges in the built environment and public capital markets (Robinson & McIntosh, 2022; Morri et al., 2024). From an environmental perspective, REITs are closely exposed to physical and transition risks (Contat et al., 2024), while their tenant relations, workforce practices, and reputation are affected by social aspects (Robinson & McIntosh, 2022; Zheng & Lee, 2026). Governance remains important due to the REIT-specific regulatory structure, agency issues, and dependence on debt capital (Zheng & Lee, 2026).

Prior ESG research in the REIT context has focused primarily on quantitative performance indicators, while also documenting that ESG effects are not uniform across all three dimensions (Aroul et al., 2022; Fan et al., 2024; Morri et al., 2024; Zheng & Lee, 2026). Separately, the disclosure literature has shown that the tone of REIT corporate filings carries value-relevant information (e.g., Carstens & Freybote, 2021; Doran et al., 2012; Koelbl, 2020). Yet the specific role of ESG-related sentiment in REIT disclosures remains unexplored, even though evidence from the broader finance literature indicates that ESG-related textual sentiment affects stock market outcomes (Dorfleitner & Zhang, 2024; Ignatov, 2023; Jeong et al., 2025; Yu et al., 2023). In particular, it is unclear whether investors respond differently to ESG-related sentiment in mandatory financial disclosures and voluntary standalone ESG reports, whether aggregate ESG sentiment is sufficient, and whether these reactions vary across REIT subsectors.

To address this gap, the aim of this paper is to determine the relevance of ESG-related sentiment in REIT corporate communication by comparing disclosure channels, examining ESG-pillar sentiment, and assessing differences across REIT subsectors.

Methodology and Data

This study analyzes mandatory corporate disclosures (10-K and 10-Q filings) as well as voluntary ESG reports of 174 U.S. REITs that were constituents of the Nareit index between 2016 and 2025. In total, the dataset comprises approximately 1,500 10-K filings, 3,600 10-Q filings, and 400 ESG reports. We first identify ESG-related text by applying a task-specific fine-tuned Robustly Optimized BERT Pretraining Approach (RoBERTa) model that classifies disclosure paragraphs into E, S, and G, while also allowing for neutral or multiple assignments. In a second step, sentiment is classified as positive, negative, or neutral at the paragraph level using a RoBERTa model fine-tuned on financial text, as no ESG-specific sentiment training dataset of comparable scale was

available. Based on the resulting classifications, we construct a Net Positive Tone (NPT) measure that captures the relative predominance of positive over negative language.

To investigate investor reactions, we estimate ordinary least squares (OLS) panel regressions in an event study setting that relates the NPT to cumulative abnormal returns (CARs) over multiple event windows around publication dates. This analysis is extended along three dimensions, comparing disclosure channels (mandatory filings, voluntary ESG reports), individual ESG pillars, and REIT subsectors, while controlling for firm-level, market-based, and macroeconomic variables to improve the robustness of our results.

Results and Their Contribution to Science and Practice

The results support a disaggregated view of ESG communication in REIT capital markets. Aggregated sentiment from ESG-related paragraphs has limited explanatory power for abnormal returns. In contrast, differentiating by disclosure channel, ESG pillar, and REIT subsector reveals systematic investor reactions. Within mandatory filings, sentiment regarding environmental aspects emerges as the most stable and economically relevant driver of abnormal returns and, in line with expectations, appears especially relevant for property types where energy efficiency and environmental risks have a direct effect on operating performance. Social sentiment shows less consistent effects across sectors, and governance sentiment is largely insignificant, likely because REIT-specific regulatory structures already mitigate these risks. In voluntary ESG reports, pricing effects remain limited, with only social-pillar sentiment exhibiting a negative association with abnormal returns.

This study extends the REIT ESG literature by shifting from quantitative ESG indicators to ESG communication in textual corporate disclosures, providing first evidence on whether and how ESG-related textual sentiment is incorporated by investors in the REIT sector. For investors and analysts, this means that ESG communication is most informative when it is assessed selectively rather than as a general ESG signal. For REIT managers, the results indicate that generic positive language in ESG communication does not generate favorable market responses. Instead, credible and economically grounded ESG disclosures are more informative to capital market participants, suggesting that disclosure strategy should prioritize substance over tone.

Paper 4: News Sentiment and Stock Return Predictability: The Case of U.S. REITs

Problems and Objective

Predicting stock market returns remains an ongoing challenge, as conventional predictive variables have shown limited out-of-sample performance (Welch & Goyal, 2008). For the REIT market, existing evidence is similarly discouraging at the index level (Cabrera et al., 2026; Ling et al., 2000).

The theoretical foundation for sentiment-driven mispricing rests on Baker and Wurgler (2006, 2007), who argue that the effect of sentiment is more pronounced for assets that are hard to value and difficult to arbitrage. REITs exhibit both characteristics through their hybrid nature. Beyond this asset-level perspective, the literature further suggests that sentiment primarily predicts the cross-section of stock returns rather than aggregate market movements (Stambaugh et al., 2012). This cross-sectional perspective is particularly relevant for real estate and REIT markets, where heterogeneity is especially pronounced across subsectors, with meaningful differences in their sensitivity to sentiment and exposure to uncertainty (Demiralay & Kilincarslan, 2024; Marcato & Nanda, 2016). However, whether sentiment carries predictive information at the cross-sectional level, where theory suggests its effects to be strongest, has not yet been examined for the REIT market in an out-of-sample setting.

The objective of this paper is therefore to investigate whether news-based sentiment exhibits out-of-sample predictive power for U.S. REIT subsector returns and whether the resulting signals can be translated into economically meaningful sector allocation decisions.

Methodology and Data

This study uses 820 weekly observations from January 2010 to September 2025. The U.S. REIT market and its subsectors are proxied by the FTSE EPRA Nareit Total Return indices, covering the Equity REITs, Office, Industrial, Residential, and Retail subindices. Additionally, capital market and macroeconomic control variables are included in our model.

The textual dataset comprises approximately 321,000 news articles from three sources: Macroeconomic & Policy News, as well as REIT News obtained from S&P Market Intelligence, and CoStar News. For sentiment classification, we compare a fine-tuned Robustly Optimized BERT Pretraining Approach (RoBERTa) model and a Generative Pre-trained Transformer (GPT) model using zero-shot prompting. To mitigate the potential look-ahead bias of GPT models, all date expressions within articles are masked, and articles for which the model correctly recalled the publication week are excluded from the analysis.

The forecasting framework follows a strict out-of-sample design with the first prediction week being the first full trading week of 2015. We employ a least absolute shrinkage and selection operator (LASSO) model as a linear benchmark and an eXtreme Gradient Boosting (XGBoost) model as a nonlinear counterpart, with annually retuned hyperparameters via time series cross-validation. The prediction task is framed in three settings: return-level forecasting as a regression, and two binary classification tasks predicting return direction and sector excess return relative to the Equity REITs index.

We evaluate model performance using standard metrics, including out-of-sample R^2 and balanced accuracy. In addition, economic relevance is assessed through a sector rotation strategy, and the SHapley Additive exPlanations (SHAP) framework is used to examine feature importance.

Results and Their Contribution to Science and Practice

Consistent with prior literature, aggregate return prediction at the weekly horizon remains elusive. However, the results suggest that sentiment contains economically meaningful information for predicting relative sector outperformance. Sentiment signals should therefore not be seen as a predictor of the overall market but as a cross-sectional allocation signal identifying which subsector may outperform others. Results differ strongly across subsectors. While the office and industrial sectors, which are more exposed to macroeconomic and capital market dynamics, show the strongest forecasting results, the residential and retail sectors remain close to random. This pattern is consistent with theoretical expectations of heterogeneous sensitivity to sentiment across asset types.

An analysis of SHAP values confirms that sentiment metrics rank among the most important model features, exceeding the contribution of traditional macroeconomic and capital market indicators, indicating that sentiment operates as a distinct predictive factor rather than merely proxying information already contained in standard data. In contrast to expectations, the sentiment derived using the fine-tuned RoBERTa model exhibits substantially more predictive power compared to that derived from the GPT model, suggesting that task alignment and proper fine-tuning outperform generative language models relying on zero-shot prompting for sentiment analysis. The economic relevance of the sentiment signal is further supported by sector rotation strategies based on the predictions of the XGBoost model with RoBERTa sentiment, which outperform both the Equity REITs index and a non-sentiment benchmark model. The implied break-even transaction costs of 12–22 basis points suggest that sentiment-driven mispricing at the sector level is economically meaningful but constrained by transaction costs at high frequencies.

For institutional investors, news-based sentiment is therefore most informative as a complement to existing quantitative models in cross-sectional investment decisions, where it captures information that is not reflected in conventional predictors. For researchers and practitioners using large language models (LLMs) for sentiment analysis in specialized markets, this finding suggests that domain adaptation through fine-tuning remains essential, even as the general capabilities of generative models continue to improve.

6.2 Final Remarks

Understanding the non-fundamental beliefs that influence asset prices has been a central research objective in behavioral finance for decades, but only recent advances in computational methods and data availability have made it possible to measure these beliefs accurately and at scale through textual sources. With the empirical relevance of sentiment for asset pricing well established, the question has shifted from whether sentiment matters to how it should be measured and through which channels it operates to provide informative signals.

The theoretical mechanisms explaining how sentiment-induced mispricing can arise and persist in markets that are traditionally assumed to be rational are rooted in the concept of noise (Black, 1986), the limits that constrain arbitrage as a corrective force (De Long et al., 1990; Shleifer & Vishny, 1997), and the idea that such mispricing is most pronounced for assets that are difficult to value and difficult to arbitrage (Baker & Wurgler, 2006, 2007). Taking these theoretical mechanisms to the REIT setting, this dissertation demonstrates that social media sentiment serves as a leading indicator for REIT returns, preceding market movements by several months (Paper 1). It further shows that the informative value of sentiment is not static but shifts from being primarily non-fundamental in normal market phases to reflecting fundamental information during periods of heightened investor engagement (Paper 2). A third finding concerns ESG-related sentiment in corporate disclosure, which is found to be processed selectively by investors, with aggregate ESG sentiment showing limited explanatory power, while disaggregation by pillar and subsector reveals systematic short-term market reactions (Paper 3). Beyond these explanatory insights, this dissertation examines whether such information can be translated into economically meaningful forecasts in an out-of-sample setting (Paper 4). The results suggest that while news-based sentiment does not predict aggregate REIT returns, it effectively identifies relative sector outperformance. Although operating in a stylized setting that does not account for all frictions of real-world implementation, the proposed sector rotation strategies support these findings and provide further evidence that textual sentiment can move from a descriptive metric to an actionable signal for asset allocation.

Taken together, these results show that textual sentiment in REIT markets provides empirical access to the non-fundamental beliefs that behavioral theory identifies as one source of persistent mispricing. A central implication is that the channel through which sentiment is observed is part of the economic interpretation of the signal. Social media, corporate disclosure, and financial news are therefore not interchangeable metrics of the same underlying sentiment but capture different phases in the formation, shaping, and processing of beliefs.

These contributions are not without boundaries. Several limitations follow from the empirical setting of the underlying studies and from the scope of sentiment as a measure. A first constraint

concerns the availability and structure of textual data. Despite the comparatively transparent U.S. REIT market, research remains limited by restricted access to textual corpora, irregular publication frequencies, and discontinuities in data coverage. This limits the construction of long, continuous textual time series necessary for in-depth analysis and complicates the direct comparison of channels at uniform frequencies. A more general constraint is that purely explanatory findings in empirical asset pricing research are inherently backward-looking. Relying on historical data does not automatically allow for generalization about future market trends, as changing economic or regulatory conditions can fundamentally alter previously observed patterns.

In addition, the cumulative structure of this dissertation imposes a limitation of its own. While three distinct textual sources are investigated, the underlying papers use different research designs, time periods, and methodologies. That sentiment differs across channels is therefore inferred from comparisons of study results, rather than from a direct comparison within a unified framework. The patterns identified in this dissertation should accordingly be read as complementary evidence from distinct analyses, not as controlled comparisons. Future research could directly address this structural limitation by comparing sentiment across information channels using the same classifiers, time periods, and empirical framework. Such a design would allow for the quantification of how much information is embedded in each channel and to what extent these signals overlap. It would also provide the basis for examining whether combining sentiment from multiple channels improves out-of-sample prediction beyond what individual channels achieve. If different channels capture distinct investor groups and stages of the belief-formation process, their joint inclusion may yield incremental predictive power. Whether this incremental information translates into measurable economic gains through improved allocation decisions naturally extends, for instance, the analysis of Paper 4.

A more fundamental boundary concerns what sentiment, as a measure, can capture. Reducing text to its evaluative direction makes large-scale empirical testing possible but leaves aside central informative dimensions, such as the causal logic or factual substance of the underlying narrative. A negative text about refinancing conditions and a negative text about tenant demand may receive the same sentiment score yet convey entirely different information relevant to different submarkets. This is therefore not a data limitation, but a conceptual boundary of sentiment-based textual analysis. The methods used in this dissertation range from dictionary-based approaches in the tradition of Tetlock (2007) and Loughran and McDonald (2011) through supervised machine- and deep learning classifiers to fine-tuned and generative language models. Yet regardless of their sophistication, all of them ultimately compress text into a scalar sentiment value. While Paper 1 demonstrates clear gains when moving from a dictionary-based to a deep learning classifier, the comparison of a fine-tuned RoBERTa classifier and a GPT model with zero-shot prompting in Paper 4 points to diminishing returns. Consequently, greater architectural

complexity does not automatically yield stronger empirical signals when the underlying task remains polarity classification. Instead, domain-specific calibration proves more consequential than model scale. Shiller's (2017) concept of narrative economics points beyond this task boundary. It describes stories with structure, causal claims, and transmission dynamics that cannot be captured by any polarity score, however precisely estimated. Generative language models may eventually make these deeper layers of text accessible at empirical scale, enabling a shift from measuring evaluative tone to tracking how narratives form and how they gain and lose intensity over time. Whether such narrative-level information adds predictive power beyond sentiment and whether it can be integrated into real-time allocation models are questions that follow directly from the analyses presented in the preceding chapters.

This dissertation has shown that textual sentiment in REIT markets is neither uniform nor universally informative. Its results demonstrate that textual analysis can make non-fundamental beliefs empirically observable and produce signals with economic value. Whether sentiment affects asset prices is no longer the prevailing question. What remains is how it can be quantified with sufficient precision across channels and contexts, and whether new language models can move the field from measuring sentiment to understanding the narratives that produce it.

6.3 References

Aroul, R. R., Sabherwal, S., & Villupuram, S. V. (2022). ESG, operational efficiency and operational performance: Evidence from real estate investment trusts. *Managerial Finance*, *48*(8), 1206–1220. https://doi.org/10.1108/MF-12-2021-0593

Baker, M., & Wurgler, J. (2006). Investor sentiment and the cross-section of stock returns. *The Journal of Finance*, *61*(4), 1645–1680. https://doi.org/10.1111/j.1540-6261.2006.00885.x

Baker, M., & Wurgler, J. (2007). Investor sentiment in the stock market. *Journal of Economic Perspectives*, *21*(2), 129–151. https://doi.org/10.1257/jep.21.2.129

Barber, B. M., & Odean, T. (2008). All that glitters: The effect of attention and news on the buying behavior of individual and institutional investors. *The Review of Financial Studies*, *21*(2), 785–818. https://doi.org/10.1093/rfs/hhm079

Black, F. (1986). Noise. *The Journal of Finance*, *41*(3), 528–543. https://doi.org/10.1111/j.1540-6261.1986.tb04513.x

Cabrera, G., Díaz, J. D., & Hansen, E. (2026). Real estate returns and the macroeconomy: Insights from big data in the US, Canada, and the UK. *The Journal of Real Estate Finance and Economics*. Advance online publication. https://doi.org/10.1007/s11146-025-10045-4

Carstens, R., & Freybote, J. (2019). Tone in REIT financial statements and institutional investments. *Journal of Property Research*, *36*(3), 227–244. https://doi.org/10.1080/09599916.2019.1650802

Carstens, R., & Freybote, J. (2021). Can the textual tone in REIT financial statements improve the information environment for commercial real estate investors? An investigation. *Journal of Real Estate Research*, *43*(3), 335–354. https://doi.org/10.1080/08965803.2021.1985923

Contat, J., Hopkins, C., Mejia, L., & Suandi, M. (2024). When climate meets real estate: A survey of the literature. *Real Estate Economics*, *52*(3), 618–659. https://doi.org/10.1111/1540-6229.12489

Cookson, J. A., Lu, R., Mullins, W., & Niessner, M. (2024). The social signal. *Journal of Financial Economics*, *158*, Article 103870. https://doi.org/10.1016/j.jfineco.2024.103870

De Long, J. B., Shleifer, A., Summers, L. H., & Waldmann, R. J. (1990). Noise trader risk in financial markets. *Journal of Political Economy*, *98*(4), 703–738. http://www.jstor.org/stable/2937765

Demiralay, S., & Kilincarslan, E. (2024). Uncertainty measures and sector-specific REITs in a regime-switching environment. The *Journal of Real Estate Finance and Economics*, *69*, 545–584. https://doi.org/10.1007/s11146-022-09898-w

Doran, J. S., Peterson, D. R., & Price, S. M. (2012). Earnings conference call content and stock price: The case of REITs. *The Journal of Real Estate Finance and Economics*, *45*(2), 402–434. https://doi.org/10.1007/s11146-010-9266-z

Dorfleitner, G., & Zhang, R. (2024). ESG news sentiment and stock price reactions: A comprehensive investigation via BERT. *Schmalenbach Journal of Business Research*, *76*, 197–244. https://doi.org/10.1007/s41471-024-00185-3

Fan, K. Y., Shen, J., Hui, E. C. M., & Cheng, L. T. W. (2024). ESG components and equity returns: Evidence from real estate investment trusts. *International Review of Financial Analysis*, *96*(Part B), Article 103716. https://doi.org/10.1016/j.irfa.2024.103716

Gallimore, P., & Gray, A. (2002). The role of investor sentiment in property investment decisions. *Journal of Property Research*, *19*(2), 111–120. https://doi.org/10.1080/09599910110110671

Hannum, C., Arslanli, K. Y., & Kalay, A. F. (2019). Spatial analysis of Twitter sentiment and district-level housing prices. *Journal of European Real Estate Research*, *12*(2), 173–189. https://doi.org/10.1108/JERER-08-2018-0036

Hausler, J., Ruscheinsky, J., & Lang, M. (2018). News-based sentiment analysis in real estate: A machine learning approach. *Journal of Property Research*, *35*(4), 344–371. https://doi.org/10.1080/09599916.2018.1551923

Ignatov, K. (2023). When ESG talks: ESG tone of 10-K reports and its significance to stock markets. *International Review of Financial Analysis*, *89*, Article 102745. https://doi.org/10.1016/j.irfa.2023.102745

Jeong, S.-H., Han, J.-J., Jun, S., Kim, S., & Kim, J. W. (2025). Investor responses to ESG news sentiment: Exploring differential effects and industry moderation. *Corporate Social Responsibility and Environmental Management*, *32*(3), 3944–3964. https://doi.org/10.1002/csr.3163

Koelbl, M. (2020). Is the MD&A of U.S. REITs informative? A textual sentiment study. *Journal of Property Investment & Finance*, *38*(3), 181–201. https://doi.org/10.1108/JPIF-12-2019-0149

Li, T., van Dalen, J., & van Rees, P. J. (2018). More than just noise? Examining the information content of stock microblogs on financial markets. *Journal of Information Technology*, *33*(1), 50–69. https://doi.org/10.1057/s41265-016-0034-2

Ling, D. C., Naranjo, A., & Ryngaert, M. D. (2000). The predictability of equity REIT returns: Time variation and economic significance. *The Journal of Real Estate Finance and Economics*, *20*(2), 117–136. https://doi.org/10.1023/a:1007821103728

Ling, D. C., Naranjo, A., & Scheick, B. (2014). Investor sentiment, limits to arbitrage and private market returns. *Real Estate Economics*, *42*(3), 531–577. https://doi.org/10.1111/1540-6229.12037

Loughran, T., & McDonald, B. (2011). When is a liability not a liability? Textual analysis, dictionaries, and 10-Ks. *The Journal of Finance*, *66*(1), 35–65. https://doi.org/10.1111/j.1540-6261.2010.01625.x

Marcato, G., & Nanda, A. (2016). Information content and forecasting ability of sentiment indicators: Case of real estate market. *Journal of Real Estate Research*, *38*(2), 165–204. https://doi.org/10.1080/10835547.2016.12091442

Maynard, D., Bontcheva, K., & Rout, D. (2012). Challenges in developing opinion mining tools for social media. In *Proceedings of @NLP can u tag #user_generated_content?! Workshop at LREC 2012* (pp. 15–22). European Language Resources Association. http://www.lrec-conf.org/proceedings/lrec2012/workshops/21.LREC2012%20NLP4UGC%20Proceedings.pdf

Morri, G., Dipierri, A., & Colantoni, F. (2024). ESG dynamics in real estate: Temporal patterns and financial implications for REITs returns. *Journal of European Real Estate Research*, *17*(2), 263–285. https://doi.org/10.1108/JERER-01-2024-0005

Ozik, G., Sadka, R., & Shen, S. (2021). Flattening the illiquidity curve: Retail trading during the COVID-19 lockdown. *Journal of Financial and Quantitative Analysis*, *56*(7), 2356–2388. https://doi.org/10.1017/S0022109021000387

Paulus, N. M., Lautenschlaeger, L., & Schaefers, W. (2025). Social media and real estate: Do Twitter users predict REIT performance? *Journal of Real Estate Research*, *47*(3), 322–355. https://doi.org/10.1080/08965803.2024.2340915

Robinson, S., & McIntosh, M. G. (2022). A literature review of environmental, social, and governance (ESG) in commercial real estate. *Journal of Real Estate Literature*, *30*(1–2), 54–67. https://doi.org/10.1080/09277544.2022.2106639

Ruscheinsky, J. R., Lang, M., & Schäfers, W. (2018). Real estate media sentiment through textual analysis. *Journal of Property Investment & Finance*, *36*(5), 410–428. https://doi.org/10.1108/JPIF-07-2017-0050

Shiller, R. J. (2017). Narrative economics. *American Economic Review*, *107*(4), 967–1004. https://doi.org/10.1257/aer.107.4.967

Shleifer, A., & Vishny, R. W. (1997). The limits of arbitrage. *The Journal of Finance*, *52*(1), 35–55. https://doi.org/10.1111/j.1540-6261.1997.tb03807.x

Sprenger, T. O., Sandner, P. G., Tumasjan, A., & Welpe, I. M. (2014). News or noise? Using Twitter to identify and understand company-specific news flow. *Journal of Business Finance & Accounting*, *41*(7 & 8), 791–830. https://doi.org/10.1111/jbfa.12086

Stambaugh, R. F., Yu, J., & Yuan, Y. (2012). The short of it: Investor sentiment and anomalies. *Journal of Financial Economics*, *104*(2), 288–302. https://doi.org/10.1016/j.jfineco.2011.12.001

Tetlock, P. C. (2007). Giving content to investor sentiment: The role of media in the stock market. *The Journal of Finance*, *62*(3), 1139–1168. https://doi.org/10.1111/j.1540-6261.2007.01232.x

Welch, I. (2022). The wisdom of the Robinhood crowd. *The Journal of Finance*, *77*(3), 1489–1527. https://doi.org/10.1111/jofi.13128

Welch, I., & Goyal, A. (2008). A comprehensive look at the empirical performance of equity premium prediction. *The Review of Financial Studies*, *21*(4), 1455–1508. https://doi.org/10.1093/rfs/hhm014

Yu, H., Liang, C., Liu, Z., & Wang, H. (2023). News-based ESG sentiment and stock price crash risk. *International Review of Financial Analysis*, *88*, Article 102646. https://doi.org/10.1016/j.irfa.2023.102646

Zamani, M., & Schwartz, H. A. (2017). Using Twitter language to predict the real estate market. In M. Lapata, P. Blunsom, & A. Koller (Eds.), *Proceedings of the 15th Conference of the European Chapter of the Association for Computational Linguistics: Volume 2, Short Papers* (pp. 28–33). Association for Computational Linguistics. https://doi.org/10.18653/v1/E17-2005

Zheng, Z., & Lee, C. L. (2026). A meta-analysis of ESG factors in the real estate investment trusts sector: Exploring their impacts on REITs performance. *Engineering, Construction and Architectural Management*, *33*(3), 2314–2334. https://doi.org/10.1108/ECAM-05-2024-0652