INSTITUTO
UNIVERSITÁRIO
DE LISBOA

Sentiment Analysis: The Influence of Social Media on the Financial Market

Willian Felipe Santa Ana

Master's in *Data Science*

Supervisor(s):

Professor Ph.D. Diana E. Aldea Mendes,
Associate Professor, Iscte - Instituto Universitário de Lisboa

October, 2025

BUSINESS SCHOOL

TECHNOLOGY AND ARCHITECTURE

Department of Quantitative Methods for Management and Economics
Department of Information Science and Technology

Sentiment Analysis: The Influence of Social Media on the Financial Market

Willian Felipe Santa Ana

Master's in *Data Science*

Supervisor(s):

Professor Ph.D. Diana E. Aldea Mendes,
Associate Professor, Iscte - Instituto Universitário de Lisboa

October, 2025

Acknowledgments

The completion of this dissertation represents the culmination of a long journey, which would not have been possible without the support of essential people.

I would like to express my most sincere gratitude to my supervisor, Professor Diana Elisabeta Aldea Mendes, PhD. The brilliance of her lectures inspired me and provided the foundational knowledge to conceive the idea and interest for this topic. Her patience, availability, and warm welcome were fundamental to the development and conclusion of this work.

I extend my thanks to the members of the examination committee (the jury), for the time dedicated to reading and analyzing this dissertation, and for the valuable feedback and suggestions that enriched and improved the final quality of this document.

A special and profound thank you to my wife, Ivanice. Her emotional and motivational support was tireless. In difficult moments, it was her strength that helped me not to lose focus and to bring this work to its completion.

Abstract

The proliferation of social media has profoundly impacted financial markets, making sentiment analysis a highly efficient tool for understanding market dynamics. This dissertation investigates the influence of social sentiment, specifically from *top influencers* on the Stocktwits platform, on the daily returns of a diverse set of financial assets, including stocks, cryptocurrencies, and Exchange-Traded Funds. Using time series analysis techniques such as the Cross Correlation Function and Granger Causality tests, along with Linear regression modeling via Ordinary Least Squares and Quantile Regression, this study quantifies both the average and conditional impact of sentiment.

Granger causality tests reveal limited predictive power of sentiment for returns (significant for NVIDIA Corporation, Alphabet Inc.), but indicate reactive sentiment patterns for others (Palantir Technologies Inc., Bitcoin). Although Linear regression via Ordinary Least Squares consistently shows a positive and statistically significant average impact of sentiment across all assets, Quantile Regression uncovers substantial heterogeneity. Three distinct patterns emerge: (1) A *Buy the Dip* effect (e.g., NVIDIA Corporation, SoFi Technologies Inc.), where positive sentiment exerts a stronger influence during market downturns (lower quantiles); (2) A *Fear Of Missing Out/Momentum* effect (e.g., NIO Inc., Ripple), where the impact of sentiment intensifies during market upturns (upper quantiles); and (3) A *Constant-Linear* effect (e.g., Meta Platforms Inc., Bitcoin), where the sentiment's impact is stable across the return distribution, validating the Ordinary Least Squares findings for these specific assets.

These results highlight the limitations of mean-based models and demonstrate that the influence of social sentiment is asset-specific and market-state dependent, varying significantly with market conditions. The discovered asymmetries offer valuable insights for investors adjusting sentiment-based strategies and for regulators monitoring market stability, particularly concerning assets prone to social media-amplified *momentum* dynamics.

Keywords: *Sentiment Analysis, Social Media, Financial Markets, Market Efficiency, Behavioral Finance*

Resumo

A proliferação das redes sociais impactou profundamente os mercados financeiros, tornando a análise de sentimento uma ferramenta altamente eficiente para a compreensão das dinâmicas de mercado. Esta dissertação investiga a influência do sentimento social, especificamente de *top influencers* na plataforma *Stocktwits*, nos retornos diários de um conjunto diversificado de ativos financeiros, incluindo ações, criptomoedas e *Exchange-Traded Funds*. Utilizando técnicas de análise de séries temporais como a Função de Correlação Cruzada e testes de Causalidade de Granger, juntamente com modelação de regressão Linear via Mínimos Quadrados Ordinários e Regressão Quantílica, este estudo quantifica tanto o impacto médio como o condicional do sentimento.

Os testes de causalidade de Granger revelam um poder preditivo limitado do sentimento para os retornos (significativo para *NVIDIA Corporation*, *Alphabet Inc.*), mas indicam padrões de sentimento reativo para outros (*Palantir Technologies Inc.*, *Bitcoin*). Embora a regressão Linear via Mínimos Quadrados Ordinários mostre consistentemente um impacto médio positivo e estatisticamente significativo do sentimento em todos os ativos, a Regressão Quantílica descobre uma heterogeneidade substancial. Emergem três padrões distintos: (1) Um efeito *Buy the Dip* (ex., *NVIDIA Corporation*, *SoFi Technologies Inc.*), onde o sentimento positivo exerce uma influência mais forte durante as quedas de mercado (quantis inferiores); (2) Um efeito *Fear Of Missing Out/Momentum* (ex., *NIO Inc.*, *Ripple*), onde o impacto do sentimento se intensifica durante as subidas de mercado (quantis superiores); e (3) Um efeito *Constante-Linear* (ex., *Meta Platforms, Inc.*, *Bitcoin*), onde o impacto do sentimento é estável ao longo da distribuição dos retornos, validando os achados do Mínimos Quadrados Ordinários para estes ativos específicos.

Estes resultados destacam as limitações dos modelos baseados na média e demonstram que a influência do sentimento social é específica do ativo e dependente do estado do mercado, variando significativamente com as condições deste. As assimetrias descobertas oferecem perceções valiosas para investidores que ajustam estratégias baseadas em sentimento e para reguladores que monitorizam a estabilidade do mercado, particularmente no que concerne a ativos propensos a dinâmicas de *momentum* amplificadas pelas redes sociais.

Palavras-Chave: *Análise de Sentimento, Redes Sociais, Mercados Financeiros, Eficiência de Mercado, Finanças Comportamentais*

Contents

List of Figures

List of Tables

List of Acronyms

BERT: Bidirectional Encoder Representations from Transformers

CLS: Classification Token

SEP: Separator Token

CRISP-DM: Cross-Industry Standard Process for Data Mining

AIC: Akaike Information Criterion

BIC: Bayesian Information Criterion

OLS: Ordinary Least Squares

blue chips: Shares of large, financially stable companies with consistent performance

EMH: Efficient-market hypothesis

CCF: Cross Correlation Function

QR: Quantile Regression

FOMO: Fear Of Missing Out

ETFs: Exchange-Traded Funds

PRISMA: Preferred Reporting Items for Systematic Reviews and Meta-Analyses

FGI: Fear and Greed Index

API: Application Programming Interface

IP: Internet Protocol

RoBERTa: Robustly Optimized BERT Pretraining Approach

NLP: Natural Language Processing

DOI: Digital Object Identifier

INTC: Intel Corporation

NIO: NIO Inc.

PLTR: Palantir Technologies Inc.

SOFI: SoFi Technologies Inc.

LCID: Lucid Group, Inc.

XRP-USD: Ripple to U.S. Dollar

XRP: Ripple

AMD: Advanced Micro Devices, Inc.

META: Meta Platforms, Inc.

NKE: NIKE, Inc.

NVDA: NVIDIA Corporation

GOOGL: Alphabet Inc.

UBER: Uber Technologies, Inc.

AMZN: Amazon.com, Inc.

PFE: Pfizer Inc.

AAPL: Apple Inc.

MSFT: Microsoft Corporation

DIA: SPDR Dow Jones Industrial Average ETF Trust

BTC-USD: Bitcoin to U.S. Dollar

CHAPTER 1

Introduction

The emergence and proliferation of social media have radically transformed how information is disseminated and consumed in contemporary society. In the financial domain, platforms such as Twitter, Reddit, and, specifically, Stocktwits, have become vibrant arenas where investors, analysts, and enthusiasts share opinions, news, and forecasts about financial assets. This constant flow of information, laden with collective sentiment, raises fundamental questions about its influence on market dynamics. Does the "noise" from social media contain information relevant to prices? Can the sentiment expressed online anticipate or even drive market movements?

Academic literature has been exploring these questions, often within the scope of Behavioral Finance and the Efficient Market Hypothesis. While classical theory suggests that prices reflect all available information, more recent approaches acknowledge the role of heuristics, cognitive biases, and collective emotions in investors' decision-making (Shiller, 2003). In this context, sentiment extracted from social media emerges as a potentially useful *proxy* for the emotional state and expectations of the market (Antweiler & Frank, 2004).

However, the mere aggregation of general sentiment may conceal more complex dynamics. Not all voices on social media carry the same weight. *Influencers*, users with many followers and high activity, may have a disproportionate impact on the formation of public opinion and, consequently, on market behavior (Sabherwal et al., 2011). Additionally, the impact of sentiment may be nonlinear; its effect could vary significantly depending on market conditions—for example, being more pronounced during periods of high volatility, euphoria, or panic.

This dissertation is situated within this context, seeking to deepen the understanding of the relationship between social sentiment, specifically that of influential users, and the performance of financial markets, with a particular focus on the heterogeneity and asymmetry of this impact.

1.1. Research Problem and Questions

Existing research on sentiment analysis in financial markets, although growing, presents gaps. Many studies focus on aggregate sentiment without differentiating the impact of different types of users. Furthermore, most regression analyses use linear models (such as Ordinary Least Squares (OLS)) that estimate only the average effect, potentially masking important nonlinear or asymmetric relationships crucial for understanding the phenomenon.

Given these limitations, the central problem of this research is to determine whether, and how, the sentiment expressed by *top influencers* on the Stocktwits platform influences the daily returns of a diverse set of financial assets, explicitly considering the possibility of asymmetric effects across different market regimes.

The main research questions guiding this study are:

(1) What is the *average* impact of Stocktwits' *top influencers'* sentiment on the daily returns of selected stocks, cryptocurrencies, and Exchange-Traded Funds (ETFs)?
(2) Is there evidence of causality relationships (in the Granger sense) between *top influencers'* sentiment and asset returns, and in which direction (prediction or reaction)?
(3) Does the impact of *top influencers'* sentiment on returns vary according to the quantile of the return distribution (i.e., is it different on days of extreme decline vs. days of extreme increase)?
(4) Are there distinct patterns in how sentiment influences different asset classes or specific assets?

1.2. Objectives

To address these questions, the following specific objectives were defined:

- Collect and process data on posts and users from the Stocktwits platform, along with market data (prices and daily returns) for 55 financial assets.
- Develop and apply a metric to identify and classify *top influencers* based on their activity and reach on the platform.
- Aggregate the sentiment ('Bullish' vs 'Bearish') expressed by *top influencers* on a daily basis for each asset.
- Analyze the temporal relationships between the sentiment and return series using the Cross Correlation Function (CCF) and Granger Causality tests.
- Estimate the average impact of sentiment on returns through linear regression with OLS.
- Investigate the heterogeneity of the sentiment impact under different market conditions through the application of Quantile Regression (QR).
- Compare the results of OLS and QR to evaluate the presence of asymmetries and identify distinct behavioral patterns among the assets.
- Interpret the results obtained in light of financial and behavioral theory, discussing their economic implications.

1.3. Summary Methodology and Data

This study adopts a quantitative approach, combining time series analysis and econometrics techniques. The primary data comes from the Stocktwits platform, covering posts and user information collected and stored in a PostgreSQL database. Market data (adjusted closing prices) is obtained for the same period through Yahoo Finance (n.d.) for a relevant period compatible with the Stocktwits data.

The analysis focuses on the metric of *average top influencer sentiment*, which represents the daily average of sentiment (-1 for *Bearish*, $+1$ for *Bullish*) from posts made by users classified as influential. Daily returns are calculated as the logarithmic percentage change of the adjusted closing prices.

The statistical tools include the CCF for visualizing correlations with *lags*, Granger Causality tests to assess temporal precedence, Linear regression via OLS for the average effect, and Quantile Regression to model the effect at multiple points of the conditional distribution of returns. The analyses were implemented in Python, utilizing libraries such as *Pandas*, *Statsmodels*, and *Scipy*.

1.4. Expected Contributions

This dissertation aims to contribute to the existing literature in several ways:

- **Focus on Influencers:** By isolating the sentiment of *top influencers*, provide insights into the specific role of this potentially more impactful group, in contrast to aggregate sentiment.
- **Application of Quantile Regression:** Explicitly demonstrate the limitations of OLS and the superiority of QR in capturing asymmetric and market-state-dependent effects in the sentiment-return relationship.
- **Identification of Heterogeneous Patterns:** Empirically document the existence of distinct patterns ("Buy the Dip," "Fear Of Missing Out (FOMO)/Momentum," "Constant-Linear") in how sentiment influences different assets, enriching the understanding of the interaction between investor psychology and asset characteristics.
- **Multi-Asset Analysis:** By including stocks from different sectors, cryptocurrencies, and ETFs, allow for a broader comparison of sensitivity to sentiment across asset classes.

1.5. Dissertation Structure

This document is organized as follows: **Chapter 2** presents a review of the relevant literature on sentiment analysis, behavioral finance, market efficiency, and the role of social media and influencers. **Chapter 3** addresses the collection and processing of data, and the construction of the dataset used in this study. **Chapter 4** covers the sentiment analysis and classification stage, along with the choice of the classification model used. **Chapter 5** details the econometric models used (CCF, Granger, OLS, QR) for the study of correlation between market values and investor sentiment. **Chapter 6** presents the empirical results of the descriptive analyses, the causality tests, and the Linear regression models by OLS estimator and Quantile regression, the graphical visualizations, including the interpretation of the results, discussing the economic and behavioral implications of the identified patterns. Finally, **Chapter 7** presents the main conclusions, acknowledges the limitations of the study, and suggests directions for future research.

CHAPTER 2

Literature Review

The increasing complexity of contemporary financial markets, driven by digitalization and the proliferation of information, has intensified interest in understanding investor behavior. Historically, financial theory was based on the premise of the rationality of economic agents, as established by the Efficient-market hypothesis (EMH) of Fama (1970). However, the observation of market anomalies and the influence of psychological factors spurred the development of behavioral finance, which recognizes the role of emotions and cognitive biases in investment decisions Yadav and Chakraborty (2023).

Within this framework, the emergence of social media and other digital platforms has substantially transformed how investors access and process information. The rapid dissemination of news, opinions, and sentiment through channels such as Twitter (currently X), online forums, and blogs exerts a significant impact on market dynamics Li et al. (2023). Investor sentiment, susceptible to the influence of these information flows, constitutes a factor that can cause asset prices to deviate from their fundamental values, creating opportunities or risks for market participants (Fan et al., 2023).

This literature review chapter examines the intersection between investor behavior, market sentiment, and the impact of social media and other sources of digital information on financial markets. The purpose is to synthesize the main findings of recent literature, identify research gaps, and discuss the implications for investors, regulators, and researchers. The analysis covers studies that investigate how emotions and herding behavior are influenced by digital information, as well as the role of fundamental and sentiment factors in asset valuation, including cryptocurrencies.

2.1. Targets

The main objective of this literature review chapter is to provide a comprehensive and critical analysis of existing research on the influence of investor sentiment and social media on investor behavior and financial market dynamics. To achieve this general objective, the review aims to:

- Synthesize the theories and models that explain the formation and impact of investor sentiment in financial markets, with emphasis on the contributions of behavioral finance.
- Analyze the role of social media and digital platforms as sources of information and channels for sentiment dissemination, examining how they influence investment decisions and the formation of speculative bubbles or herding behavior, according to Li et al. (2023) and Cioroianu et al. (2024).

- Identify the main factors, both fundamental and sentiment-related, that affect the valuation of different asset classes, including cryptocurrencies, and how the interaction between these factors shapes market returns (Bakhtiar et al., 2023; Fan et al., 2023).
- Discuss the practical implications of the findings for investors, portfolio managers, regulators, and policymakers, offering perspectives on investment strategies, risk management, and market regulation.
- Point out gaps in the existing literature and suggest directions for future research, contributing to the advancement of knowledge in the domain of investor behavior and digital finance.

2.2. Research Methodology

The proliferation of social media as a space for public debate has created new challenges and opportunities in the context of the financial market. Sentiment analysis, defined as the use of computational techniques to interpret emotions and opinions expressed in digital media, has been widely applied to predict trends in asset appreciation and depreciation (Baker & Wurgler, 2006; Tetlock, 2007). This study utilizes a systematic literature review to explore the relationship between sentiments expressed in social media and investor behavior. For this purpose, relevant articles were selected using the query ((“social media sentiment”) OR (“web media analysis”)) AND ((“stock market”) OR (“stock exchange”)) AND (“investor behavior”) in the Scopus and Web of Science databases.

Initially, 183 records were identified, of which 83 came from the Scopus database and 100 from the Web of Science. The identification and selection of articles followed the Preferred Reporting Items for Systematic Reviews and Meta-Analyses (PRISMA) protocol, adapted from Page et al. (2021), which we illustrate in Figure 2.1. After applying initial filters, restricting the results to journal articles, published in English, and with a publication date from 2021 onwards, 117 records were selected. Among these, four duplicates were removed, resulting in 114 records for detailed title and abstract analysis. At this stage, 80 studies were excluded for not meeting the relevance criteria, such as the lack of focus on investor behavior or sentiment analysis. Thus, 33 articles remained for full-text evaluation.

During the eligibility phase, 11 additional articles were excluded, nine for not directly addressing investor behavior and two for not having accessible full text. Based on the defined criteria, 22 studies were finally included in this systematic review. The rationale for the literature selection was based on the need to capture recent studies that address contemporary advances in understanding investor behavior in digital contexts. Priority was given to works that analyzed sentiments extracted from digital platforms and their influence on investment decisions, considering the increasing use of data from social media as indicators of investor behavior and perceptions.

Additionally, this systematic review considers the impact of different types of digital networks addressed in the selected articles, which we will discuss later. Among these,

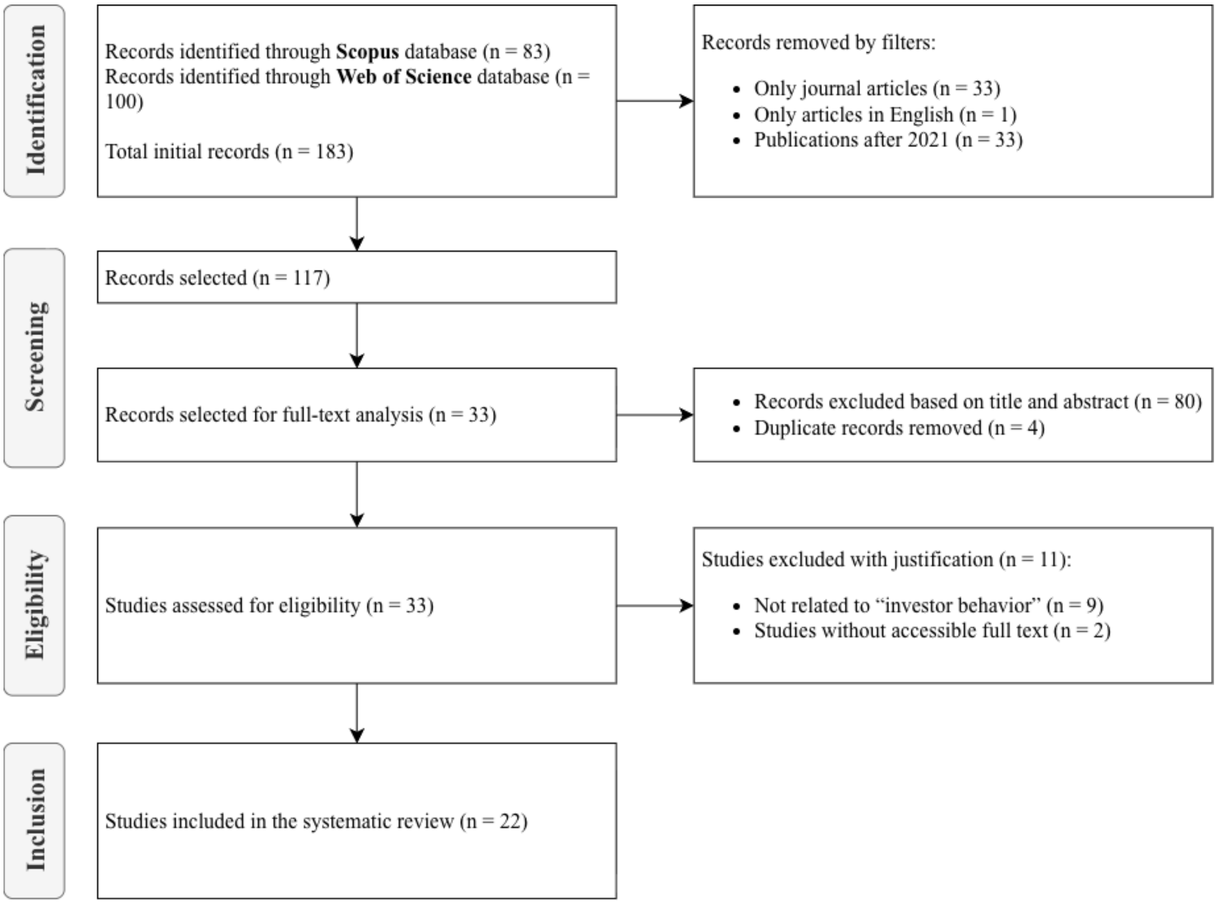

FIGURE 2.1. Flow diagram of the research using the PRISMA protocol.

social networks such as Twitter and Reddit, financial news platforms, specialized blogs, and online discussion forums, such as Eastmoney, stand out. Each type of network plays a distinct role in the formation and dissemination of investor sentiment. For example, studies highlight the role of Twitter in identifying changes in market sentiment (Chen et al., 2014; Fan et al., 2023) and the influence of forums like Eastmoney on the interaction between individual investors, which can both promote the dissemination of information and amplify market noise (Wu et al., 2024).

2.2.1. Analysis of Selected Studies and Academic Journals

The analysis of the studies selected for the literature review was organized based on the number of citations, as well as the quality of the academic journals in which they were published. Table 2.1 presents the articles ordered by the number of citations. It is observed that most articles have few citations, which reflects the recent nature of the studies included in this review. This fact is understandable, given that many of these works were published after 2021, which limits the time available for their dissemination and impact within the scientific community. Despite this, articles such as *Introspecting predictability of market fear in Indian context during COVID-19 pandemic*, with 35 citations, and *The Effect of Online Environmental News on Green Industry Stocks: The Mediating Role of Investor*

Sentiment, with 20 citations, demonstrate relevance and broader acceptance within the field.

Despite the reduced quantitative impact, these works address important issues, contributing to the understanding of specific topics, such as the use of computational models to analyze financial bubbles driven by social media. The low citation count in this context should be interpreted in light of the emerging nature of research in the domain of investor behavior and sentiment analysis in digital networks.

TABLE 2.1. Papers identified in the systematic review obtained on November 5, 2024.

Rank	Article Title	Citations
1	Interrogating predictability of market fear in Indian context during COVID-19 pandemic	35
2	The Effect of Online Environmental News on Green Luxury Stocks: The Mediating Role of Investor Sentiment	20
3	Firm-level tweet sentiment and corporate announcement returns	13
4	Understanding the role of social media sentiment in identifying irrational herding behavior in the stock market	12
5	Social informativeness and investor sentiment in the GubaStock post message	12
6	Stock-level sentiment alongside and the cross-section of stock returns	8
7	The Impact of Different Sentiment in Investment Decisions: Evidence from China's Stock Markets IPOs	5
8	Shadow of words and monetary printing	4
9	The influence of investor sentiment and sentiment on the valuation of cryptocurrencies	4
10	Integrating fundamentals into investor behavior	3
11	Positive Price Walk Behavior Based on an Explanatory Multitasking Based Analytical Model for Financial Investments	2
12	Predicting Prices from Investor Sentiment Analysis: A Study of S&P's, Stealth&II and Grunefi Cryptocurrencies	2
13	Social Media Stock Sentiment Sequence Construction	2
14	The Influence of Social Media on Investment Decisions-Making: Examining Behavioral Biases, Risk Perception, and Modulation Effect	2
15	Sentiment Announcements and Stock Market Returns in India	2
16	Aggregate sentiment versus disaggregated social interaction and the information efficiency of the Chinese capital market	2
17	An Optimizers' strategies with traditional vs. digital assets and geopolitics and banking	2
18	Incorporating social media information into Chinese open-end fund investment learning for portfolio optimization	2
19	Measuring Diversified Media Information Impact on the Chinese Peer-to-Peer Lending Market	2
20	The role of aggregating investor sentiment across cryptocurrencies	2
21	Sentiment of Social Media-Driven Bubble Formation in Financial Markets using an Agent-Based Model	2
22	Unveiling COVID-19's impact on Financial Stability	2

Table 2.2 organizes the journals where the articles were published, based on data from Scimago Journal & Country Rank (n.d.). The classification considers the journal quartile and citations per publication in 2023. The results show that most studies were

published in high-impact journals classified in Quartile 1, such as the *International Journal of Information Management Data Insights*, which obtained 15.526 citations per publication on average, and *Blockchain: Research and Applications*, with 10.54 citations per publication. These journals stand out for their academic rigor and relevance in the field of financial sciences and information management.

TABLE 2.2. List of Journals ordered by quartile and citations/publications for the year 2023. According to SCImago (n.d.) obtained on November 5, 2024.

Rank	Journal	Total	Citations / Pub. (2 years)	Quartile
1	International Journal of Information Management Data Insights	2127	19.224	Q1
2	Blockchain: Research and Applications	696	10.504	Q1
3	Electronic Markets	4276	9.376	Q1
4	Research in International Business and Finance	4878	7.198	Q1
5	Research in International Business and Finance	4276	6.646	Q1
6	International Review of Economics and Finance	3879	5.966	Q1
7	International Review of Economics and Finance	3379	5.646	Q1
8	Net Access	208752	5.476	Q1
9	North American Journal of Economics and Finance	2041	5.148	Q1
10	Journal of Banking and Finance	3011	5.004	Q1
11	Humanities and Social Sciences Communications	4946	4.946	Q1
12	Heliyon	43116	4.916	Q1
13	Economic Research-Ekonomska Istrazivanja	4444	4.444	Q1
14	Physica A: Statistical Mechanics and its Applications	10779	4.051	Q2
15	International Journal of System Assurance Engineering and Management	2071	3.977	Q2
16	Journal of Risk and Financial Management	1997	3.931	Q2
17	Journal of Futures Markets	496	3.828	Q2
18	Finance Research Letters	3542	3.654	Q2
19	Sustainability	11628	3.102	Q2
20	International Journal of Economics and Financial Issues	107	2.085	Q3
21	Contemporary Mathematics (Springer)	35	1.000	Q4
22	European Modeling and Simulation Symposium, EMSS	91	0.664	Q4

2.3. Results and Discussion

2.3.1. The Role of Investor Sentiment and Social Media

Recent literature has consistently demonstrated that investor sentiment, often amplified by social media, plays a significant role in asset price formation and market behavior. Bird et al. (2023) provide robust evidence that emotions, as measured in news and social media, influence investment decisions and asset valuation. They observe that positive emotions, such as confidence and optimism, tend to have a more pronounced impact on investor reactions than negative emotions. Interestingly, the study also points out that emotions based on traditional news still exert a greater influence on stock valuations than those originating from social media, although the influence of the latter is rapidly growing over time.

> *"We find strong evidence to support that emotions do influence investor decisionmaking and provide important insights into the nature of this relationship. In general, we find those positive emotions such as trust and optimism are more influential in shaping investors' reactions than are negative emotions. Finally, the emotions based on the news media listings have a greater influence on stock valuations than those based on social media listings." (Bird et al., 2023, p. 1)*

However, the relationship between social media and investor behavior is not linear and can lead to ambiguous results. Wu et al. (2024), by analyzing the Chinese capital market, identified a U-shaped relationship between online investor interaction and stock price synchronicity. This suggests that, at moderate levels, online communication can disseminate relevant information and promote the "wisdom of the crowds." However, beyond a certain threshold, the increase in interactions contributes to market noise, distorting the price discovery process and stimulating irrational emotions that harm market efficiency. This finding is essential as it highlights the need to discern between useful information and the mere "hype" generated on social platforms.

> *"Our findings indicate that online investor communication can effectively disseminate firm-specific information through social networks when investor interactions are at a moderate level. However, beyond a certain threshold, heightened interactions contribute to market noise, thereby distorting the stock price discovery process." (Wu et al., 2024, p. 1)*

2.3.2. Sentiment and Specific Assets: The Case of Cryptocurrencies

The application of sentiment analysis is particularly relevant in emerging and volatile markets, such as cryptocurrencies, where traditional fundamentals may be less defined and investor sentiment plays a more prominent role. Bakhtiar et al. (2023) investigated the impact of fundamental and sentiment factors on cryptocurrency valuation. The study, published in Blockchain: Research and Applications, a journal noted for its focus on distributed ledger technologies and their financial applications, revealed that the Fear and Greed Index (FGI) and Google Trends are significant factors affecting cryptocurrency values. The journal's connection to the topic is direct, as it addresses the underlying technology and market dynamics of cryptocurrencies, an asset strongly influenced by sentiment.

> *"The research shows the importance of sentiments and suggests that the fear and greed index can indicate when to make a cryptocurrency investment, while Google search interest in cryptocurrency is crucial when*

choosing the appropriate type of cryptocurrency." (Bakhtiar et al., 2023, p. 1)

This study deepens the understanding of how sentiment, captured by indicators like FGI and Google Trends, can predict market movements in an environment where information is often asymmetric and speculation is high. The relevance of Google Trends, for example, suggests that public interest and attention can drive speculative bubbles, a critical point for investors and regulators. The critical analysis here lies in recognizing that, while these sentiment indicators are powerful, they can also signal the formation of bubbles, requiring caution from investors.

2.3.3. Herding Behavior and Irrationality

Herding behavior, where investors imitate each other's actions without considering their own information or analysis, is a common manifestation of sentiment's influence. Li et al. (2023) explore the role of social media sentiment in identifying irrational herding behavior in the Chinese stock market. They conclude that social media sentiment has a significant impact on irrational herding behavior, and that the authority and influence of social media accounts can have a differential moderating effect. Specifically, official accounts tend to be more neutral, while posts from unofficial accounts with many followers are more likely to influence irrational herding behavior. This finding is fundamental to understanding market dynamics, as it differentiates between the "wisdom of the crowds" and the "irrational herd," which can lead to suboptimal decisions.

> *"It is found that the impact of social media sentiment in firm-specific information on stock market investor sentiment is significant and exacerbates irrational stock market herding behavior; furthermore, the authority and influence of social media accounts can have a differential impact on irrational herding behavior." (Li et al., 2023, p. 1)*

2.3.4. Implications and Challenges for Research

The findings of the reviewed literature have significant implications for various stakeholders. For investors, understanding the impact of sentiment and social media can lead to more informed investment strategies that consider not only fundamentals but also market psychology. For regulators, the need to monitor the flow of information on social media and develop policies that mitigate the dissemination of misleading information and irrational herding behavior becomes evident. The International Review of Economics and Finance, where Li et al. (2023)'s article was published, is an important venue for discussions on economic and financial policies, which reinforces the relevance of such studies for formulating effective regulations. Despite the advances, significant gaps still exist in the literature. Most studies focus on developed markets, and there is a scarcity of in-depth research on the sentiment-return relationship in developing economies, according to Yadav and Chakraborty (2023). Furthermore, the clear distinction between the effects of the rational and irrational components of investor sentiment on stock performance is not yet fully consensual. Future research could explore the application of machine learning and artificial intelligence techniques to identify and quantify sentiment in a more granular way, as well as investigate the interaction between different types of sentiment (e.g., fear, optimism, anxiety) and their impact on various asset classes and market conditions. The analysis of how macroeconomic and geopolitical events influence sentiment and, consequently, investor behavior in different cultural contexts also represents a promising area for future studies.

It has been observed that sentiment, captured by various metrics, has a tangible impact on market returns, especially in volatile environments like cryptocurrencies. The ability of social media to rapidly disseminate information and sentiment can both promote market efficiency through the "wisdom of the crowds" and generate irrationality and herding behavior. The distinction between these two effects is fundamental for investors and regulators.

Although the field has advanced significantly, there is still vast ground to be explored. The need for more research in emerging markets, the elucidation of the interaction between different types of behavioral biases, and the development of more sophisticated methodologies to capture and analyze sentiment are promising areas. The continuous evolution of technology and social communication platforms will ensure that the intersection between finance, psychology, and technology remains a fertile field for academic research and of great practical relevance.

CHAPTER 3

Data Collection

This chapter describes the methodology employed in the development and execution of this dissertation, detailing the approaches for data collection and processing. The objective is to ensure the replicability and validity of the obtained results, focusing on the integration of sentiment analysis from social platform messages with financial market data.

3.1. Methodological Framework: CRISP-DM

The Cross-Industry Standard Process for Data Mining (CRISP-DM) model serves as this study's methodological foundation. CRISP-DM is a widely recognized and utilized approach in data science projects, offering a structured and iterative lifecycle for managing *data mining* projects (Chapman et al., 2000). This model is composed of six main phases, which will be adapted and detailed in the context of this research:

- **Business Understanding**: This initial phase focuses on understanding the project objectives from a business perspective, defining success criteria and data requirements. Within this dissertation, the main objective is to investigate the influence of social sentiment, expressed on financial platforms, on the behavior of the financial market.
- **Data Understanding**: Involves the initial data collection, verification of its quality, exploration of the data to gain preliminary insights, and the formulation of hypotheses. This phase is essential for familiarizing the researcher with the available data sources and their characteristics.
- **Data Preparation**: Covers all activities necessary to construct the final dataset from the raw data. This includes data selection, cleaning, construction, and integration. This is an intensive phase that can consume a significant portion of the project time.
- **Modeling**: In this phase, appropriate modeling techniques are selected and applied to the prepared data. This includes algorithm selection, model construction, and its validation. Although the detailed analysis of the models is addressed in a later chapter, the sentiment classification methodology, which is a prerequisite for modeling, will be detailed here.
- **Evaluation**: The constructed model is evaluated against the business objectives, verifying whether the results are satisfactory and if the model is robust. This phase may lead to a return to previous phases for refinement.

- **Deployment**: Consists of putting the model into practical use. In the context of an academic dissertation, the deployment can be the presentation of results and the provision of a prototype or *framework* for future research.

The following sections detail the Data Understanding and Data Preparation phases, which are the main focus of this methodological chapter, as well as the sentiment classification methodology, which is a critical component of the modeling phase for this study.

3.2. Data Understanding and Collection

Data collection was carried out from two distinct sources: a social platform specialized in financial discussions (Stocktwits, 2025) and historical financial market data sources (Yahoo Finance, n.d.). The integration of these sources allowed for the construction of a comprehensive dataset for subsequent analysis.

3.2.1. Financial Social Platform Data

The social platform was selected as the primary source of sentiment data due to its specialized nature in financial discussions and the use of specific identifiers (e.g., Apple Inc. (AAPL), Bitcoin to U.S. Dollar (BTC-USD)) that directly associate messages with financial assets. Data collection was performed through an automated process of interaction with the platform, simulating user behavior to access and extract the necessary information. This process was implemented using browser automation libraries, such as Playwright, which allowed for navigation, authentication, and interaction with dynamic page elements (e.g., infinite scrolling) that would not be accessible through static extraction methods. This method was adopted due to the absence of a public Application Programming Interface (API) that would allow for the direct collection of a high volume of data.

The data collected for each message includes:

- **Message Identification**: A unique identifier for each post.
- **Associated Asset**: The identifier of the financial asset to which the message refers.
- **Message Author**: The username of the post's author.
- **Date and Time of Publication**: The exact temporal record of the publication.
- **Textual Content**: The full text of the message.
- **Interactions**: The number of comments, shares, and likes that the message received. These metrics are dynamic and were updated periodically to reflect the most recent state of interactions.
- **Multimedia Content**: If the message contained an image, it was downloaded and processed. Images were resized for space and loading optimization, while maintaining their original aspect ratio.
- **Sentiment**: A sentiment classification associated with the message, obtained through a specialized sentiment analysis model, detailed in section 3.4.

To optimize collection efficiency, a parallel processing architecture was adopted, allowing multiple workers to collect data from different assets or authors simultaneously. This approach was essential for building the dataset within a viable timeframe. Data persistence was achieved in a relational database, ensuring the integrity and query capability of the collected data.

The continuous extraction process presented significant challenges, such as managing user sessions, Internet Protocol (IP) blocks, and handling dynamically loaded content. To ensure the robustness and integrity of the data, several strategies were implemented: (1) a detailed logging system in the database; (2) duplicate checking before insertion to ensure collection idempotency; and (3) automatic worker restart mechanisms to prevent failures due to memory exhaustion or session loss.

3.2.2. Financial Market Data

Historical market data for the assets corresponding to the identifiers collected on the social platform were obtained through publicly accessible financial data sources. These sources provide detailed information on asset performance over time, including opening, high, low, closing, adjusted closing prices, and daily trading volume (Yahoo Finance, n.d.).

For each asset, additional market metrics, essential for subsequent analysis, were calculated:

- **Daily Return**: Calculated as the percentage change in the daily closing price. For a day t and closing price P_t, the return R_t is given by:

  $$R_t = \frac{P_t - P_{t-1}}{P_{t-1}}$$

  This metric is fundamental for evaluating the asset's daily performance and is widely used in financial market studies (Campbell et al., 1997).
- **Volatility**: Measured as the standard deviation of daily returns within a rolling window. This metric reflects the dispersion of returns and, consequently, the asset's risk over a short period. Volatility is a key indicator of market uncertainty and price fluctuation (Engle, 1982).

3.3. Data Preparation and Enrichment

After the initial collection, the data underwent a preparation and enrichment process, essential to ensure the quality and suitability of the data for subsequent analysis phases. This process included updating interaction metrics, collecting author data, and refactoring historical data.

3.3.1. Updating Interaction Metrics

Interaction metrics (comments, shares, etc.) were not always reliably loaded in the main feed of an asset, which was the target of the initial collection. Therefore, a secondary process was implemented that visited the individual page of each message whose interaction counter was null, thus ensuring the accurate and complete collection of this data.

3.3.2. Author Data Collection and Enrichment

The enrichment of author data involved two distinct approaches. First, the public profile of each author was extracted to obtain static influence metrics, such as the number of followers and followed users. Second, average engagement metrics (average of likes, shares, and comments) were calculated based on the history of each author's posts already present in our dataset, allowing for an assessment of their actual influence within the collected sample, enabling the subsequent segmentation of authors based on their level of impact on the community (Bakshy et al., 2011).

3.3.3. Refactoring Historical Data

This step involved migrating data from a legacy database structure (a table of tweets from previous exploratory collections) to the final schema. During this process, the associated multimedia content (images in Base64 format) was decoded, reprocessed for optimization (resizing and compression), and stored as static files, ensuring consistency with the newly collected data, and allowing for modularity with the separation of images from the final dataset.

3.4. Dataset Availability

The dataset resulting from this collection and preparation process, which forms the empirical basis of this dissertation, was made openly available to the scientific community on the Zenodo platform (CERN & OpenAIRE, n.d.). Zenodo is an open-access, general-purpose repository, developed under the OpenAIRE program and operated by CERN, which allows for the sharing of research data, software, and other digital materials, ensuring their long-term preservation and the assignment of a Digital Object Identifier (DOI) for easy citation and traceability.

This availability aims to promote transparency, the replicability of results, and to facilitate future research in the field of sentiment analysis in financial markets. The final dataset, named *Sentiment Analysis in Financial Social Media: Impact on Market Volatility* (Santa Ana, 2025), is composed of:

- **953,209 messages** collected from the Stocktwits social platform.
- **55 different symbols**, covering stocks, cryptocurrencies, and Exchange-Traded Funds (ETFs).
- **57,470 different authors** with associated influence and engagement data.

The publication of this resource demonstrates a commitment to the principles of Open Science, transforming the work of this phase into an asset that adds value for the advancement of new research involving this topic.

CHAPTER 4

Sentiment Analysis and Classification

Sentiment analysis constitutes a fundamental component of this research, requiring a well-defined methodological approach in the selection of the classification model. The choice of an appropriate model for sentiment analysis in financial contexts presents unique challenges that differ significantly from generic sentiment analysis. The language used in financial communications is characterized by specific jargon, technical acronyms, references to market events, and semantic nuances that may not be adequately captured by models trained on generic data (Loughran & McDonald, 2011).

To ensure the accuracy and relevance of the sentiment classification, a comparative evaluation of four specialized sentiment analysis models was performed, with particular emphasis on models adapted for the financial domain. This evaluation was based on a benchmark dataset that allows for an objective comparison of the performance of the candidate models.

4.1. Evaluated Models

To ensure the selection of the most effective model, a comparative *benchmark* was conducted. Four notable sentiment analysis models were evaluated on a public reference dataset, `zeroshot/twitter-financial-news-sentiment`, which contains 9,543 financial messages labeled as negative (0), positive (1), or neutral (2).

The models selected for evaluation include:

- **FinTwitBERT-sentiment** (Akkerman, 2023): A specialized adaptation of the Bidirectional Encoder Representations from Transformers (BERT) model specifically fine-tuned for sentiment analysis on financial data from social media. This model was trained on an extensive corpus of financial messages, making it particularly suitable for capturing the nuances of financial language on platforms like Stocktwits.
- **BERTweet** (FiniteAutomata, 2021): A model based on the BERT architecture, optimized for sentiment analysis on social media messages, with a particular focus on the characteristic structure and language of Twitter and similar platforms.
- **Twitter-RoBERTa** (CardiffNLP, 2023): An implementation of the Robustly Optimized BERT Pretraining Approach (RoBERTa) model adapted for sentiment analysis on social media content, offering improvements in the pretraining architecture compared to the original BERT.

- **FinBERT** (ProsusAI, 2020): A BERT model specialized for sentiment analysis in financial texts, trained on formal financial documents and corporate communications, representing a more traditional approach to financial sentiment analysis.

4.1.1. Theoretical Foundation of the BERT Architecture

The BERT model, which forms the basis of the evaluated models, represents a significant innovation in the field of natural language processing. The BERT architecture is based on the *Transformers* mechanism (Vaswani et al., 2017), which uses attention mechanisms to process text sequences bidirectionally.

The *self-attention* mechanism constitutes the core of the Transformer architecture, allowing the model to weigh the importance of different parts of the input sequence when processing each token. This mechanism can be expressed mathematically as:

$$\text{Attention}(Q, K, V) = \text{softmax}\left(\frac{QK^T}{\sqrt{d_k}}\right)V$$

Where Q (Query), K (Key), and V (Value) are matrices derived from the input *embeddings*, and d_k is the dimension of the *keys* vectors. This formulation allows the model to capture long-range dependencies and understand the complete context of the message.

The bidirectional capability of BERT is particularly advantageous for sentiment analysis, as it allows the model to consider both the preceding and succeeding context of each word, resulting in a deeper understanding of the sentiment expressed in the message.

4.1.2. Evaluation Methodology

The evaluation of the models was conducted using a reference dataset composed of 9,543 previously classified financial messages, distributed across three sentiment categories: negative (1,442 messages), neutral (6,178 messages), and positive (1,923 messages). This distribution reflects the natural prevalence of different types of sentiment in financial communications, with a predominance of neutral content.

For each model, standardized performance metrics were calculated, including *accuracy*, weighted F_1-score, and F_1-score per class. The F_1-score is particularly relevant in this context, as it combines precision and *recall* into a single metric, defined as:

$$F_1 = 2 \times \frac{\text{Precision} \times \text{Recall}}{\text{Precision} + \text{Recall}}$$

Where precision represents the proportion of correct classifications among all positive classifications for a class, and *recall* represents the proportion of correctly identified positive instances.

4.2. Comparative Evaluation Results

The evaluation results reveal significant differences in the performance of the candidate models, as shown in Table 4.1.

TABLE 4.1. Performance of candidate models in financial sentiment classification

Model	Accuracy	Weighted F1-Score	F1 Negative	F1 Neutral	F1 Positive
FinTwitBERT	0.8340	0.8387	0.7817	0.8765	0.7600
BERTweet	0.6709	0.6658	0.5537	0.7566	0.4583
Twitter-RoBERTa	0.6901	0.6823	0.4917	0.7775	0.5194
FinBERT	0.7121	0.7202	0.6088	0.7863	0.5911

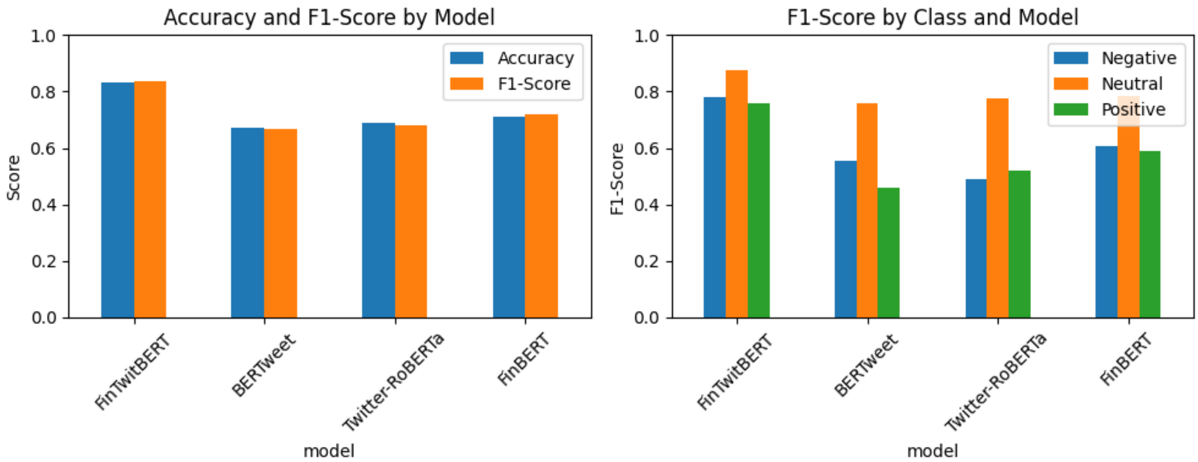

FIGURE 4.1. Accuracy and F1-Score by model - F1-Score by class and by model

The **FinTwitBERT-sentiment** model demonstrated the best overall performance, achieving an accuracy of 83.40% and a weighted F1-score of 0.8387. This superior result can be attributed to its specialized training on financial social media data, which allows it to more effectively capture the specific nuances of the language used on the Stocktwits platform.

The analysis of the confusion matrices (Figure 4.2) also reveals that *FinTwitBERT* exhibits less confusion between sentiment classes, demonstrating a superior ability to distinguish between positive, negative, and neutral sentiments. Given the results, the **FinTwitBERT-sentiment** model was chosen to classify the sentiment of the data collected from Stocktwits, as it will help ensure higher quality sentiment classification for use in subsequent analyses.

4.3. Sentiment Classification Process

Before applying the classification model, the messages collected from the *Stocktwits* platform were subjected to a cleaning and normalization process optimized to preserve information relevant to financial sentiment analysis. This process includes:

- **Removal of Irrelevant Elements**: Elimination of user mentions (`@username`) and URLs, which do not directly contribute to the message's sentiment but can introduce noise into the classification.

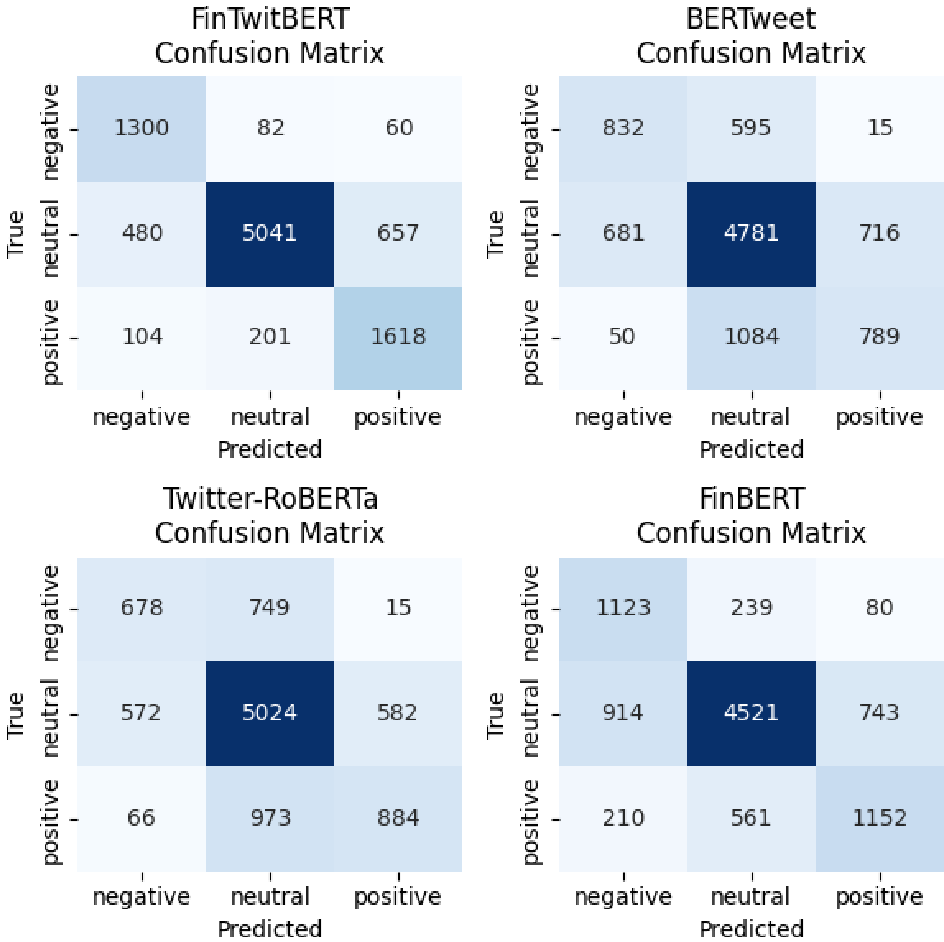

FIGURE 4.2. Confusion matrices for the models

- **Preservation of Financial Symbols**: Maintenance of *cashtags* (`$SYMBOL`) and *hashtags* (`#tag`), which constitute indispensable informational elements in financial communications and can significantly influence the expressed sentiment.
- **Textual Normalization**: Conversion to lowercase and removal of unnecessary punctuation, while maintaining the essential semantic structure of the message.

4.3.1. Classification Pipeline

The sentiment classification process followed a structured *pipeline* that maximizes processing accuracy and efficiency:

- **Tokenization**: The cleaned text is divided into *tokens* (words or subwords) using the *tokenizer* associated with the *FinTwitBERT* model. This process includes the addition of special *tokens*, (Classification Token (CLS) and Separator Token (SEP)), necessary for the correct functioning of the BERT model.
- **Encoding**: The *tokens* are converted into numerical representations (*embeddings*) that capture semantic and contextual meaning. The model uses *token*, segment, and positional *embeddings* to create a rich representation of the input.

- **Transformer Processing**: The *embeddings* are processed through the multiple layers of the *Transformer* model, where each layer applies *self-attention* mechanisms and *feed-forward* neural networks to refine the contextual representation.
- **Final Classification**: The representation of the CLS *token* is fed into a linear classification layer that maps the contextualized representation to the three sentiment classes: **BULLISH** (positive), **BEARISH** (negative), and **NEUTRAL** (neutral).

4.3.2. Confidence Metrics

For each classification, the model provides a confidence score that indicates the certainty of the prediction. This metric is calculated using the *softmax* function applied to the outputs of the classification layer:

$$\text{Confidence} = \max(\text{softmax}(z_i))$$

Where z_i represents the scores (*logits*) for each class i. Confidence ranges between 0 and 1, with values close to 1 indicating high certainty in the classification.

4.4. Sentiment Classification Results

The application of the **FinTwitBERT-sentiment** model to the complete dataset resulted in the classification of 953,209 messages collected between December 2024 and March 2025. The overall sentiment distribution reveals interesting patterns in investor behavior on the *Stocktwits* platform:

TABLE 4.2. Distribution of Sentiments in Classified Messages

Sentiment Category	Number of Messages	Percentage
BULLISH (Positive)	470,724	49.4%
BEARISH (Negative)	366,271	38.4%
NEUTRAL (Neutral)	116,214	12.2%

The predominance of positive sentiment (49.4%) over negative (38.4%) suggests a general optimistic bias among platform users during the analyzed period. This distribution is consistent with previous studies that document a tendency towards optimism in financial communications on social media (Da et al., 2011).

The relatively low proportion of neutral messages (12.2%) may reflect the polarized nature of financial discussions, where investors tend to express defined opinions about the assets they are interested in.

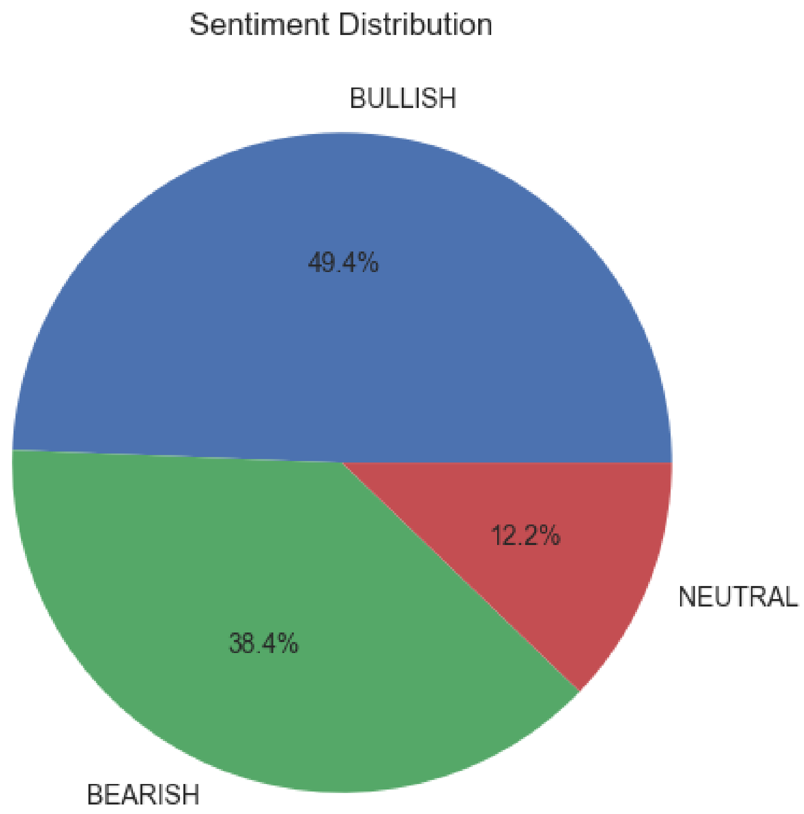

FIGURE 4.3. Distribution of Sentiments in Classified Messages

The overall average confidence of the classifications was 0.8281, indicating a high level of certainty in the model's predictions. This metric reinforces the reliability of the generated sentiment labels, which are fundamental for subsequent quantitative analyses.

The distribution of confidence by sentiment category reveals relevant variations between classes:

- **BULLISH**: Average confidence of 0.9102
- **BEARISH**: Average confidence of 0.9198
- **NEUTRAL**: Average confidence of 0.7175

It is observed that classifications associated with *bullish* and *bearish* sentiments exhibit substantially higher confidence levels than *neutral* messages. This suggests that the *FinTwitBERT-sentiment* model is more assertive in identifying content with clear polarization (positive or negative), while demonstrating greater uncertainty in detecting neutral messages—an expected behavior in sentiment analysis tasks, given the ambiguous and less emotional nature of this type of language.

Thus, although the overall average confidence is high, the discrepancy between categories reflects the intrinsically more diffuse nature of neutral messages, rather than a model deficiency. On the contrary, the high confidence observed in the polarized classes reinforces the model's robustness in capturing clearly optimistic or pessimistic opinions.

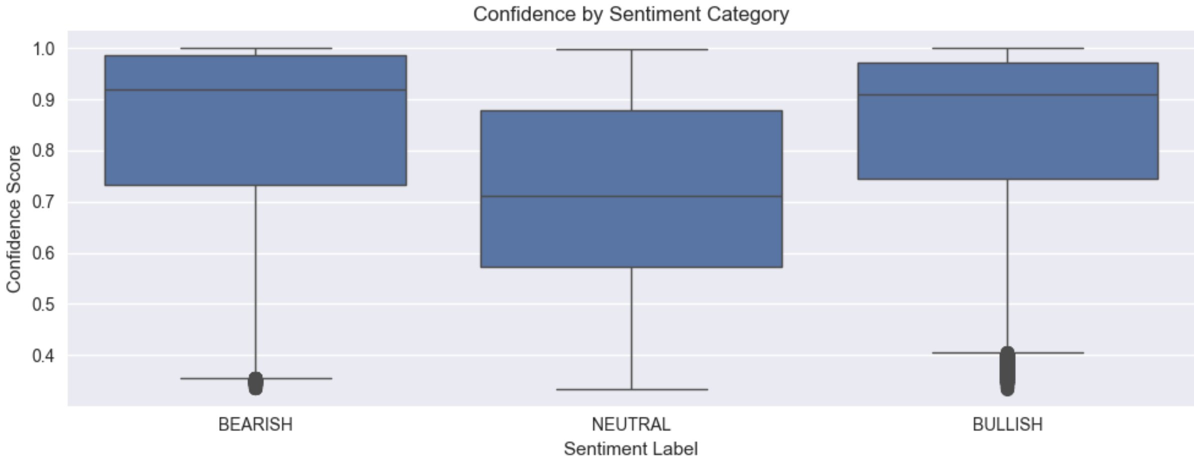

FIGURE 4.4. Distribution of Confidence by Sentiment Category

4.4.1. Temporal Evolution of Sentiments

Figure 4.5 presents the temporal evolution of the three sentiment categories—*Bullish*, *Bearish*, and *Neutral*—observed in messages posted on the *Stocktwits* platform between December 2024 and June 2025. This graph allows for the visualization of the collective dynamics of investor sentiment over the analysis period, revealing patterns of emotional behavior associated with market events and the volatility of financial assets.

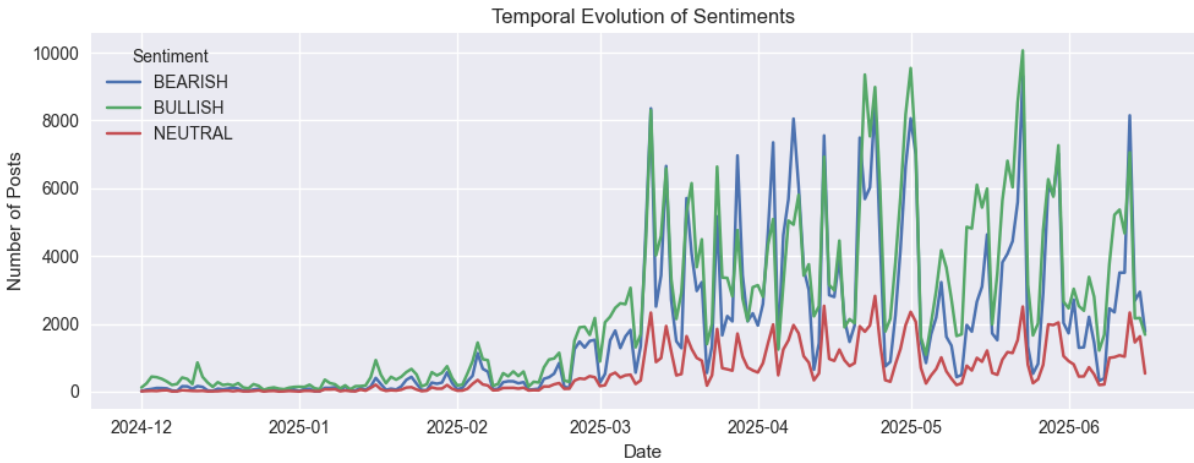

FIGURE 4.5. Temporal Evolution of Sentiments — Daily volume of messages classified as *Bearish*, *Bullish*, and *Neutral*.

A gradual increase in the volume of posts is observed starting in January 2025, with sharp peaks between March and May of the same year. This behavior coincides with a period of high volatility in international financial markets, possibly associated with the release of quarterly earnings reports from major technology companies and relevant macroeconomic events.

Bullish (optimistic) sentiment predominates at various times during the analyzed period, demonstrating a consistent positive bias among investors in the *Stocktwits* community. However, peaks in *bearish* (pessimistic) sentiment occur intermittently and tend to coincide with abrupt increases in the total volume of messages, suggesting that negativity intensifies in response to specific market events or unfavorable news.

Meanwhile, *neutral* sentiment maintains a relatively stable presence, but with less quantitative expression, reflecting the natural tendency of investors to express polarized opinions (positive or negative) in contexts of strong price variation.

Overall, the temporal evolution of sentiments demonstrates that the collective behavior of investors is highly sensitive to informational shocks and periods of uncertainty, reinforcing the relevance of sentiment analysis as a complementary indicator for volatility and return forecasting studies. This temporal pattern will be subsequently correlated with market data of selected assets to investigate the relationship between sentiment fluctuations and price movements.

4.4.2. Lexical Analysis of Negative Sentiments

Figure 4.6 presents the most frequent words in messages classified with *bearish* (negative) sentiment, obtained from the lexical analysis of the *Stocktwits* platform's corpus of posts. The objective of this analysis is to identify the main terms and expressions associated with investors' pessimistic perceptions, offering a qualitative view complementary to the quantitative sentiment classification.

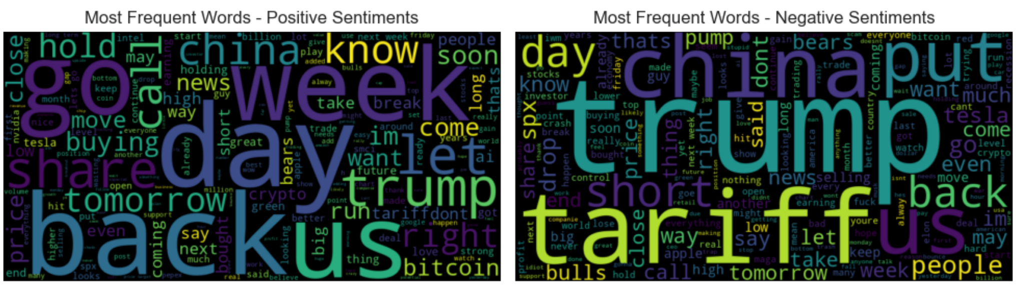

FIGURE 4.6. Most Frequent Words — Negative Sentiments

Before the construction of the word cloud and the calculation of frequencies, linguistic processing was performed to eliminate textual noise and allow the most frequent terms to effectively reflect the semantic content of the messages. This process involved the use of a custom *stopwords* list, which was expanded beyond common English words, also including:

- Financial symbols (*cashtags*) present in the dataset, such as `$AAPL`, `$TSLA`, or `$BTC`, as well as their normalized variations (e.g., `btc-usd`, `btcx`);
- Generic and recurring terms in financial discussions, but semantically neutral or redundant, such as `market`, `stock`, `today`, `money`, `buy`, `sell`, among others;
- Words directly related to the sentiment classification itself, such as `bullish` and `bearish`, to avoid circular bias in the lexical analysis.

This procedure allowed the visualized vocabulary to represent genuinely informative concepts, removing noise from repetitive mentions of assets or generic market expressions.

For positive sentiments (left graph), the most recurrent words are *"week"*, *"go"*, *"good"*, and *"trump"*. The frequency of terms like *"hold"*, *"run"*, *"big"*, and *"back"* (suggesting "recovery" or "uptrend") reflects a clear optimism and the expectation of upward movements in the market. The prominence of *"bitcoin"* highlights the interest in this asset, while the presence of *"trump"* and *"news"* suggests that positive sentiment was strongly linked to political-economic news and events during the analysis period.

On the other hand, the negative sentiment cloud (right graph) is dominated by a central theme. The most frequent word is, by far, *"trump"*, followed closely by *"tariff"* (and *"tariffs"*), *"china"*, and *"us"*. This pattern strongly indicates that investor pessimism during the analyzed period was primarily driven by political-economic uncertainties and trade tensions. Classic market terms such as *"short"*, *"down"*, *"bad"*, *"loss"*, and *"crash"* are also present, but with visibly lower frequency, suggesting they were a consequence of the main concern with politics and trade tariffs.

4.5. Considerations on Classification Quality

The quality of the sentiment classification is fundamental to the validity of the dissertation's results. Although the *FinTwitBERT-sentiment* model has been validated and is specialized, it is important to recognize that no Natural Language Processing (NLP) model is perfect. Factors such as irony, sarcasm, double meanings, and the evolution of financial slang can pose challenges. The use of a model pretrained on a relevant *corpus* minimizes these challenges, but the interpretation of the results must always consider the probabilistic nature of sentiment classification.

Previous studies have demonstrated the effectiveness of Transformer-based models for sentiment analysis in specific domains (e.g., (Araci, 2019; Yang et al., 2020)). The application of this methodology allows for the large-scale quantification of sentiment, an essential step for its integration with financial market data and the investigation of its relationships.

CHAPTER 5

Methodologies for Modeling and Correlation

This chapter deepens the analysis, modeling, and evaluation phases of the data, framed within the Cross-Industry Standard Process for Data Mining (CRISP-DM) lifecycle. The main focus lies in exploring the relationships between social sentiment, extracted from financial platforms, and financial market behavior, through the application of advanced statistical and econometric techniques. The processes of final data preparation for analysis, feature engineering, influence metrics and sentiment aggregation, correlation and causality analyses, and the applied regression models will be detailed.

5.1. Considerations on the Modeling and Evaluation Phases

In the modeling process, data analysis techniques are selected and applied to the previously prepared data. The objective is to build models that can answer the research questions, namely, if and how social sentiment influences financial markets. This phase involves variable selection, the choice of appropriate modeling algorithms, and model calibration. For this study, modeling focuses on quantifying the relationship between social sentiment and market variables, using correlation, causality, and regression approaches.

The evaluation phase is essential for determining the quality, robustness, and usefulness of the developed models. It involves interpreting the statistical results, assessing the significance of the relationships found, and discussing the implications for the financial domain. The results are compared with the initial hypotheses and the existing literature, identifying relevant *insights*, model limitations, and potential directions for future research.

5.2. Final Data Preparation for Analysis

Before applying the modeling techniques, the collected and pre-processed data were subjected to a final preparation, which included loading information from the database and integrating different data sources, as well as creating new relevant variables (*Feature Engineering*).

5.2.1. Data Loading and Integration

Message and author data, previously stored in a relational database, were loaded into tabular data structures (*DataFrames*) to facilitate manipulation and analysis. Simultaneously, historical financial market data, obtained from public sources, were integrated with this information. This integration is fundamental for aligning social sentiment with the price movements of the corresponding assets, allowing for the analysis of temporal relationships and ensuring the validity of subsequent analyses.

5.2.2. Author Segmentation

To understand the differentiated impact of sentiment, message authors were categorized based on their influence on the social platform. This categorization was carried out by calculating a composite influence score that integrates various activity and interaction metrics. Author segmentation is an important step to identify whether the sentiment of users with greater reach or engagement has a different weight on market behavior.

The metrics considered for the calculation of the influence score include:

- **Total Followers**: The number of users who follow the author, a direct indicator of the potential reach of their messages.
- **Influence Ratio**: The ratio between the number of followers and the number of users the author follows. A high ratio suggests that the author is followed more than they follow, indicating greater influence.
- **Post Volume**: The total number of messages published by the author, reflecting their activity on the platform.
- **Original Post Count (Parent)**: The number of posts that started a new discussion, as opposed to replies to other messages. This metric, inspired by works such as that of Bakshy et al. (2011), can indicate the author's ability to generate new discussion topics.
- **Average Interactions per Post**: Includes the average number of likes, shares, and comments that the author's posts receive. These metrics reflect the engagement that the author's content generates in the community.

To combine these heterogeneous metrics into a single *score*, each metric was first normalized to a common scale (i.e., between 0 and 1). Normalization can be performed using the Min–Max formula:

$$N_x = \frac{x - x_{\min}}{x_{\max} - x_{\min}}$$

where N_x is the normalized value, x is the original value of the metric, $x_{\min}$ is the minimum value of the metric in the dataset, and $x_{\max}$ is the maximum value.

After normalization, the composite influence *score* (S_I) was calculated as a weighted sum of the normalized metrics:

$$S_I = w_1 \cdot N_{\text{followers}} + w_2 \cdot N_{\text{ratio}} + w_3 \cdot N_{\text{volume}} + w_4 \cdot N_{\text{parent}} + w_5 \cdot N_{\text{interactions}}$$

Where w_i are the weights assigned to each normalized metric, reflecting its relative importance in defining influence. These weights are defined based on theoretical and empirical considerations.

Based on this *score*, authors were segmented into influence categories (e.g., *regular*, *influential*, *highly influential*, *top influencer*) using quantiles. This segmentation allows for the analysis of whether the sentiment expressed by more influential authors has a distinct impact on the market compared to less influential authors.

5.2.3. Feature Engineering

Feature engineering is the process of creating new variables from existing ones, with the aim of improving the performance of predictive models and the interpretability of the results. In this study, feature engineering focused on creating market and sentiment variables that capture the relevant dynamics for the analysis.

For market data, the following features were calculated:

- **Daily Return:** Calculated as the percentage change in the daily closing price. Formally, for a day t and closing price P_t, the return R_t is given by:

$$R_t = \frac{P_t - P_{t-1}}{P_{t-1}}$$

  This metric serves to evaluate the asset's daily performance and is widely used in financial market studies (Campbell et al., 1997).

- **Volatility:** Measured as the standard deviation of daily returns in a 5-day rolling window. This metric reflects the dispersion of returns and, consequently, the asset's risk over a short period. Volatility is an indicator of market uncertainty and price fluctuation (Engle, 1982).

For sentiment data, in addition to the aggregated metrics (see section 4.2.4), additional features could be considered, such as:

- **Sentiment Change**: The difference in average sentiment between consecutive days, to capture abrupt changes in market mood.
- **Sentiment Polarization**: A measure of the dispersion of sentiment, indicating whether opinions are concentrated at one extreme (very bullish or very bearish) or more distributed.
- **Sentiment Message Volume**: The total number of messages with positive or negative sentiment, instead of just the proportion, to capture the absolute volume of polarized discussion.

5.2.4. Sentiment Aggregation

The sentiment of the messages, previously classified (see Chapter 4), was aggregated at a daily level by asset and by author influence category. The textual sentiment ("bullish," "bearish," "neutral") was mapped to numerical values (1, -1, 0, respectively) to allow for mean and proportion calculations. The aggregated sentiment metrics include:

- **Average Sentiment:** The average of the sentiment values of the messages for a given day, asset, and influence category. It represents the general polarity of the sentiment. It is calculated as:

$$\text{Average Sentiment} = \frac{\sum_{i=1}^{N} S_i}{N}$$

  Where S_i is the sentiment value of message i and N is the total number of messages for the period and category considered.

- **Message Count:** The total number of messages for a given day, asset, and influence category. It indicates the volume of discussion and can be a proxy for market attention.
- **Bullish Proportion:** The percentage of messages with positive sentiment. It reflects the predominant optimism. Calculated as:

$$\text{Bullish Proportion} = \frac{\text{Number of Positive Messages}}{\text{Total Number of Messages}} \times 100\%$$

- **Bearish Proportion:** The percentage of messages with negative sentiment. It reflects the predominant pessimism. Calculated as:

$$\text{Bearish Proportion} = \frac{\text{Number of Negative Messages}}{\text{Total Number of Messages}} \times 100\%$$

These metrics were structured to facilitate their integration with market data, allowing for comparative analysis between different influence categories and asset behavior. Daily aggregation is a common approach in financial sentiment studies to align with the frequency of market data (Bollen et al., 2011).

5.3. Correlation and Causality Analysis

After data preparation and aggregation, correlation and causality analyses were performed to investigate the relationships between social sentiment and market variables.

5.3.1. Correlation Analysis: Pearson and Spearman

Correlation measures the strength and direction of the linear or monotonic relationship between two variables. Two main correlation measures were used to capture different aspects of the relationship:

- **Pearson Correlation Coefficient:** Measures the strength and direction of the linear relationship between two continuous variables. It is sensitive to linear relationships and assumes that the variables are approximately normally distributed. It is calculated by the formula:

$$\rho_{X,Y} = \frac{\text{cov}(X,Y)}{\sigma_X \sigma_Y} = \frac{E\big[(X-\mu_X)(Y-\mu_Y)\big]}{\sigma_X \sigma_Y}$$

  Where $\text{cov}(X,Y)$ is the covariance between X and Y, σ_X and σ_Y are the standard deviations of X and Y, and μ_X and μ_Y are the respective means. Values range from -1 (perfect negative correlation) to +1 (perfect positive correlation), with 0 indicating the absence of linear correlation (Pearson, 1895). A positive correlation indicates that the variables move in the same direction, while a negative correlation indicates they move in opposite directions.
- **Spearman Correlation Coefficient:** Measures the strength and direction of the monotonic relationship between two variables, i.e., whether one variable tends to increase or decrease when the other increases, regardless of whether the relationship is linear. It is a non-parametric measure, based on the *ranks* of the

data, making it robust to *outliers* and non-linear relationships. It is calculated by the formula:

$$\rho = 1 - \frac{6 \sum d_i^2}{n(n^2 - 1)}$$

Where d_i is the difference between the *ranks* of the observations for each pair of variables, and n is the number of observations. The Spearman coefficient also ranges from -1 to +1. It is particularly useful when the relationship between the variables is not strictly linear but follows a consistent trend (Spearman, 1904).

These analyses were applied to quantify the association between the sentiment metrics (average by influence categories, bullish and bearish for top influencers) and the market variables (returns and volatility). The comparison between the Pearson and Spearman coefficients can contribute to understanding the nature of the relationship (linear vs. monotonic).

5.3.2. Granger Causality Test

To investigate whether social sentiment can predict future asset price movements, the Granger causality test was employed (Granger, 1969). This statistical test assesses whether the past values of one time series (X, e.g., aggregate sentiment) provide statistically significant information for predicting the future values of another time series (Y, e.g., asset returns), beyond the information provided by the past values of Y (Granger, 1980).

Formally, to test the null hypothesis that X does not Granger-cause Y, two regressions are estimated:

(1) **Restricted Regression:** The variable Y is regressed only on its own past values:

$$Y_t = \alpha_0 + \sum_{i=1}^{p} \alpha_i Y_{t-i} + \epsilon_t$$

(2) **Unrestricted Regression:** The variable Y is regressed on its own past values and the past values of X:

$$Y_t = \beta_0 + \sum_{i=1}^{p} \beta_i Y_{t-i} + \sum_{j=1}^{q} \gamma_j X_{t-j} + u_t$$

The Granger test evaluates the null hypothesis that the coefficients γ_j are jointly equal to zero for all j. If the null hypothesis is rejected (i.e., if the past values of X contribute significantly to the prediction of Y), it is concluded that X Granger-causes Y. The number of *lags* (p and q) to include in the regressions is typically determined by information criteria such as Akaike Information Criterion (AIC) or Bayesian Information Criterion (BIC).

It is essential to note that Granger causality only indicates temporal precedence and does not imply a strict cause-and-effect relationship. That is, if X Granger-causes Y, it means that X contains useful information for predicting Y that is not contained in the past values of Y (Hamilton, 1994).

5.4. Regression Modeling

To quantify the impact of social sentiment on market variables, linear regression and quantile regression models were applied, allowing for an analysis of the average effects and the effects at different points of the return distribution.

5.4.1. Linear Regression with OLS Estimator

Linear regression constitutes a fundamental statistical model used to examine the relationship between a continuous dependent variable and one or more independent variables. Its objective is to model the linear relationship between these variables, allowing for the prediction of the dependent variable based on the values of the independent variables.

The multiple linear regression model can be formally expressed as:

$$Y_i = \beta_0 + \beta_1 X_{1i} + \beta_2 X_{2i} + \cdots + \beta_k X_{ki} + \epsilon_i$$

In this formulation, Y_i represents the dependent variable for the i-th observation, X_{ji} denotes the j-th independent variable for the i-th observation, β_0 is the intercept term, and β_j are the regression coefficients that quantify the marginal effect of X_j on Y, holding the other variables constant. The term ϵ_i corresponds to the random error, which encompasses unobserved factors that influence Y_i and are not captured by the independent variables included in the model.

To estimate the coefficients β of the linear regression model, the Ordinary Least Squares (OLS) method is widely employed. OLS seeks to determine the values of the coefficients that minimize the sum of the squared residuals, i.e., the difference between the observed values of the dependent variable (Y_i) and the values predicted by the model ($\hat{Y}_i$). Mathematically, OLS minimizes $\sum_{i=1}^{n} \epsilon_i^2$ (Wooldridge, 2016).

OLS is a preferred estimator due to its computational simplicity and the interpretability of its results, providing estimates of the average effect of the independent variables on the dependent variable. However, the validity of the statistical inferences derived from OLS (such as hypothesis tests and confidence intervals) depends on the satisfaction of certain classical assumptions. These include the linearity of the relationship, the independence of the errors, homoscedasticity (constant variance of the errors), and the normality of the errors. The violation of these assumptions can compromise the efficiency and impartiality of the OLS estimators, requiring the application of alternative estimation methods or corrections to ensure the robustness of the conclusions.

5.4.2. Quantile Regression

Quantile regression, introduced by Koenker and Bassett (1978), is a robust alternative to linear regression with the OLS estimator, which allows modeling the effect of the independent variables at different points of the dependent variable's distribution (quantiles), and not just at its conditional mean. This is particularly useful in financial markets, where the impact of sentiment can be asymmetric, affecting extreme returns (e.g., the tails of the distribution) more than average returns (Koenker, 2005).

The quantile regression model for the τ-th quantile of Y given X is:

$$Q_\tau(Y_i|X_i) = \beta_{0\tau} + \beta_{1\tau}X_{1i} + \cdots + \beta_{k\tau}X_{ki}$$

Where $Q_\tau(Y_i|X_i)$ is the τ-th conditional quantile of Y_i given X_i, and $\beta_{j\tau}$ are the regression coefficients for the τ-th quantile. Instead of minimizing the sum of squared residuals, quantile regression minimizes the weighted sum of absolute deviations, with different weights for positive and negative deviations, depending on the quantile τ:

$$\min_\beta \left[\sum_{i:Y_i \geq X_i^T\beta} \tau|Y_i - X_i^T\beta| + \sum_{i:Y_i < X_i^T\beta} (1-\tau)|Y_i - X_i^T\beta| \right]$$

Quantile regression is less sensitive to *outliers* and does not assume a specific distribution for the errors, making it more flexible for analyzing complex and heterogeneous relationships. It allows for the investigation, for example, of whether social sentiment has a more pronounced impact on days of large market drops or rises, providing insights that OLS would fail to capture. This technique is particularly relevant in finance, where return distributions are often non-normal and relationships can vary significantly across the distribution (Chernozhukov & Hansen, 2005).

CHAPTER 6

Analysis and Visualization of Results

This chapter is dedicated to the presentation and interpretation of the results obtained from data analysis and modeling, with a particular focus on visualization to facilitate the understanding of the identified dynamics. The findings are discussed in detail, supported by tables and graphs that illustrate the relationships between social sentiment and financial market behavior.

6.1. Causality and Regression Analysis

Causality and regression analysis allowed not only for the identification of correlations between social sentiment and market variables but also for the investigation of causal relationships and the quantification of the impact of sentiment on asset returns. This section presents the results of the correlation analysis, Granger causality tests, and regression models, providing a comprehensive understanding of the dynamics between investors' social behavior and financial market performance.

6.1.1. Correlation Analysis between Sentiment and Market Variables

Correlation analysis represents the first step in understanding the relationships between social sentiment metrics and market variables. Pearson and Spearman correlation coefficients were applied to capture both linear and monotonic relationships between the variables.

The presented correlation matrix reveals significant patterns in the relationships between social sentiment and market variables. The strongest correlations are observed between the sentiment of the most influential authors and asset returns, suggesting that the opinions of users with greater credibility have a more pronounced impact on price movements.

The results demonstrate that several technology stocks exhibit the highest correlations with social sentiment. Noteworthy are Palantir Technologies Inc. (PLTR) ($r = 0.672, p < 0.001$), NVIDIA Corporation (NVDA) ($r = 0.538, p < 0.001$), and Alphabet Inc. (GOOGL) ($r = 0.560, p < 0.001$), all showing positive and statistically significant associations between returns and influencer sentiment.

Cryptocurrencies also reveal relevant correlations, such as Ripple to U.S. Dollar (XRP-USD) ($r = 0.520, p < 0.001$) and Bitcoin to U.S. Dollar (BTC-USD) ($r = 0.379, p < 0.01$), confirming the sensitivity of these assets to social sentiment. This pattern is consistent with the literature, which points out that cryptocurrency markets are particularly susceptible to the influence of retail investors and their speculative nature (Bollen et al., 2011; Kristoufek, 2013).

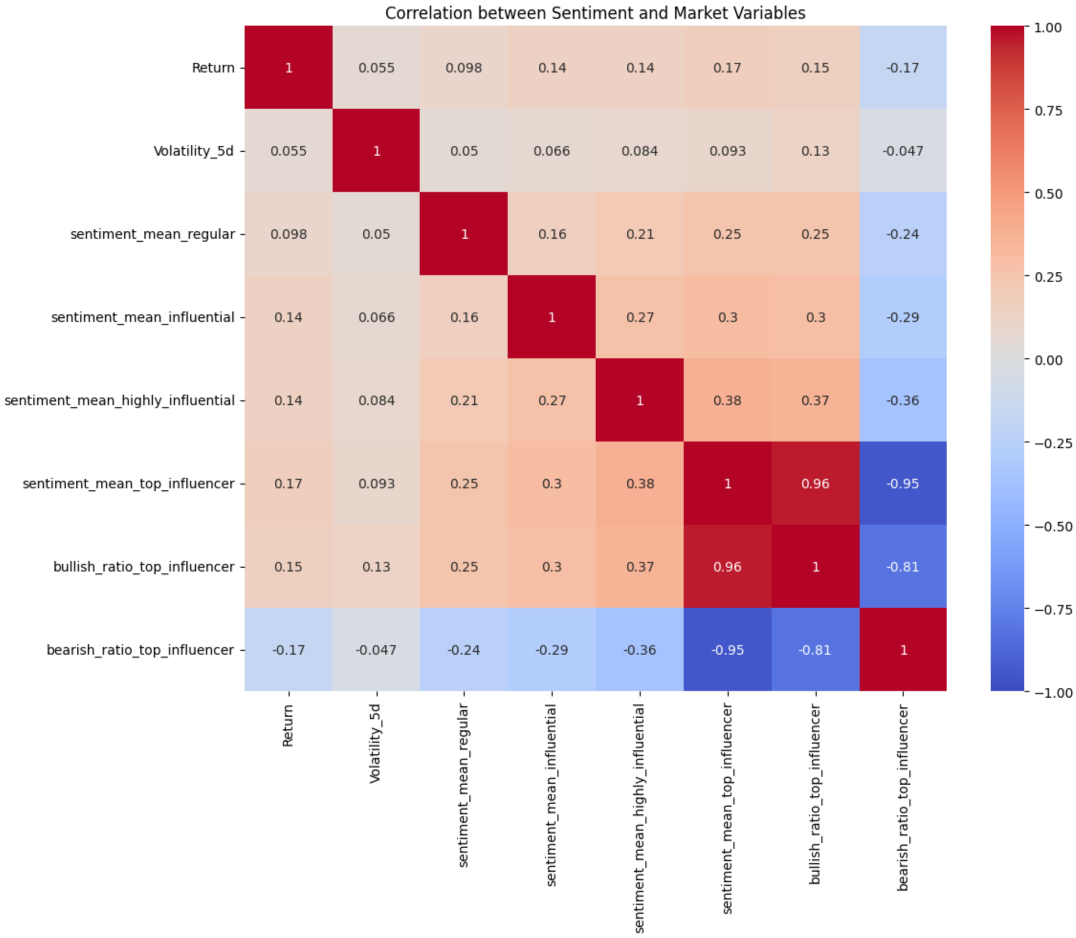

FIGURE 6.1. Correlation between Sentiment and Market Variables.

TABLE 6.1. Correlation between Return and Average Sentiment (Top Influencer)

Symbol	Pearson Corr.	Pearson p-value	Spearman Corr.	Spearman p-value
PLTR	0.671737	1.867624e-05	0.667447	2.207305e-05
GOOGL	0.559891	1.196212e-07	0.671960	2.224605e-11
INTC	0.557131	1.434552e-06	0.654808	3.276472e-09
NIO	0.607052	5.151793e-08	0.604821	5.939400e-08
LCID	0.587280	1.407618e-07	0.602873	5.322185e-08
META	0.514407	7.667880e-07	0.665872	8.737812e-12
NVDA	0.537631	4.996812e-04	0.493818	1.628440e-03
XRP-USD	0.519581	1.090342e-06	0.555128	1.330515e-07
SOFI	0.522454	7.099223e-05	0.671647	5.019239e-08
UBER	0.506107	1.233766e-07	0.525526	3.256904e-08
MSFT	0.461416	1.283096e-05	0.538510	1.803933e-07
PFE	0.465394	3.343164e-05	0.501917	6.068773e-06
NKE	0.493104	6.258096e-05	0.544851	6.768484e-06
AMZN	0.410784	1.538064e-04	0.544682	1.759166e-07
AMD	0.398418	9.221929e-04	0.503084	1.668643e-05
BTC-USD	0.379382	1.990056e-03	0.518223	1.154027e-05
DIA	0.424527	1.533708e-03	0.540177	2.982256e-05

The difference between the Pearson and Spearman coefficients is generally small, suggesting that the relationships are predominantly linear, although in some cases (e.g., SoFi Technologies Inc. (SOFI) and Meta Platforms, Inc. (META)) the Spearman correlations indicate slightly stronger monotonic relationships.

6.1.2. Cross-Correlation and Granger Causality Analysis

To investigate the temporal relationship between social sentiment and market performance, this section applies two tests: the Cross Correlation Function (CCF) with *lags* and the Granger causality test (Granger, 1969). The CCF visualizes the correlation between the two series at different *lags* (days), where negative *lags* suggest that sentiment precedes the market (predictive signal) and positive *lags* suggest that sentiment follows the market (reactive signal). A peak at the center (lag 0) indicates a contemporary correlation (on the same day).

The Granger causality test was applied to statistically evaluate these two directions:

(1) **Predictive Signal**: Whether sentiment (past lags) "Granger-causality" returns (Sentiment $\rightarrow$ Return).
(2) **Reactive Signal**: Whether returns (past lags) "Granger-causality" sentiment (Return $\rightarrow$ Sentiment).

The results are divided into two tables for each direction of causality.

TABLE 6.2. Granger Causality Test (Predictive): Sentiment $\rightarrow$ Return (p-values by lag)

Symbol	Lag 1	Lag 2	Lag 3	Lag 4	Lag 5
PLTR	0.2085	0.2774	0.0755	0.1783	0.2111
NIO	0.5999	0.8188	0.8591	0.6123	0.7660
LCID	0.7108	0.4543	0.3218	0.3805	0.1702
GOOGL	0.8774	**0.0140**	0.0689	0.0538	0.0532
INTC	0.7711	0.2715	0.5057	0.5246	0.4703
NVDA	**0.0486**	0.2546	0.1558	0.2578	0.2850
SOFI	0.9830	0.9827	0.9786	0.7912	0.8386
XRP-USD	0.4461	0.6964	0.6422	0.6284	0.4252
META	0.7333	0.9563	0.9682	0.8942	0.9367
UBER	0.8413	0.9276	0.8938	0.8153	0.8728
NKE	0.0800	0.1801	0.5036	0.3845	0.2067
PFE	0.3196	0.2701	0.4747	0.5293	0.5256
MSFT	0.7497	0.2591	0.7953	0.7105	0.7107
DIA	0.7390	0.9554	0.9732	0.9845	0.8675
AMZN	0.3695	0.4779	0.7693	0.9149	0.9385
AMD	0.6401	0.3017	0.3033	0.4665	0.7869
BTC-USD	0.8721	0.3667	0.4147	0.8020	0.8462

Note: Values in bold indicate statistical significance at the 5% level ($p < 0.05$).

TABLE 6.3. Granger Causality Test (Reactive): Return → Sentiment (p-values by lag)

Symbol	Lag 1	Lag 2	Lag 3	Lag 4	Lag 5
PLTR	0.4587	0.2152	**0.0420**	**0.0414**	**0.0432**
NIO	0.5871	0.4395	0.7019	0.5348	0.5258
LCID	0.5304	0.5501	0.7262	0.5520	0.6226
GOOGL	0.9082	0.5150	0.6797	0.5537	0.7806
INTC	0.5780	0.2746	0.5026	0.5085	0.4983
NVDA	0.7493	0.7789	0.8737	0.8594	0.8237
SOFI	0.9928	0.3173	0.5765	0.6160	0.4259
XRP-USD	0.6657	0.3351	0.5639	0.6242	0.7418
META	0.4076	0.5828	0.5314	0.8145	0.9529
UBER	0.2928	0.4356	0.3267	0.5615	0.7156
NKE	0.3959	0.6011	0.7820	0.7895	0.8990
PFE	0.7860	0.8658	0.8797	0.8061	0.7033
MSFT	0.5930	0.2005	0.3490	0.2127	0.3345
DIA	0.9917	0.8199	0.8926	0.9178	0.9610
AMZN	0.6290	0.3858	0.5757	0.4155	0.2436
AMD	0.3552	0.4793	0.7288	0.3626	0.3047
BTC-USD	**0.0101**	**0.0346**	0.5980	0.6481	0.7499

Note: Values in bold indicate statistical significance at the 5% level ($p < 0.05$).

The analysis of the causality tests reveals three distinct patterns:

(1) **Predictive Signal (Sentiment → Return):** As per Table 6.2, evidence of predictive causality was found for two assets: NVDA at lag 1 ($p = 0.0486$) and GOOGL at lag 2 ($p = 0.0140$). This suggests that the sentiment of *top influencers* today contains statistically significant information to predict the returns of NVDA one day in the future and of GOOGL two days in the future.

The graphical analysis of the CCF corroborates the direction of these findings. Figure 6.2 (NVDA) exhibits positive correlation at lag -1, although its magnitude is modest, aligning with the statistical significance found in the Granger test. Similarly, Figure 6.3 (GOOGL) shows its most prominent correlation peak at lag -2, visually indicating that sentiment precedes the market return for these two assets at the identified *lags*.

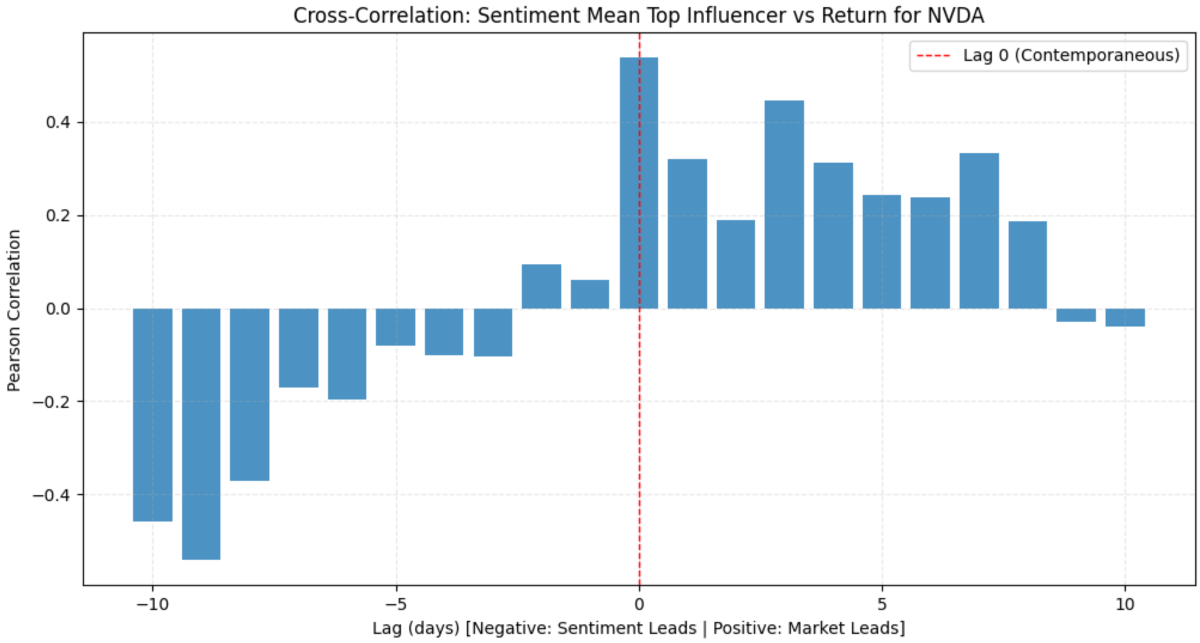

FIGURE 6.2. Cross-Correlation (NVDA): Sentiment Mean Top Influencer vs Return. (Predictive Signal at lag -1)

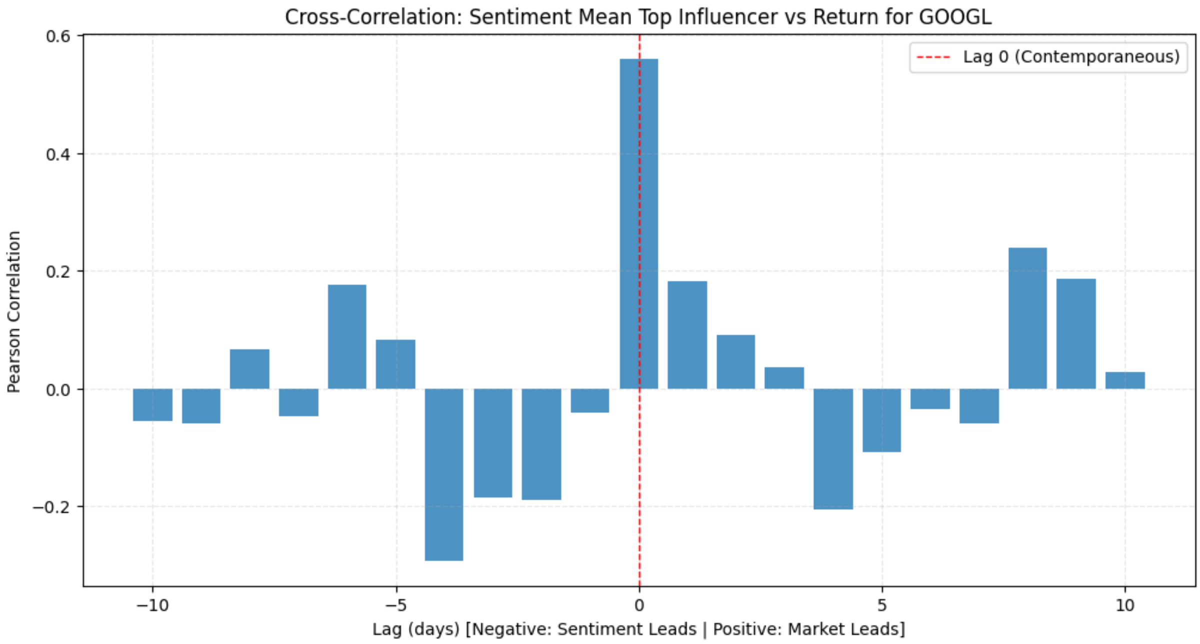

FIGURE 6.3. Cross-Correlation (GOOGL): Sentiment Mean Top Influencer vs Return. (Predictive Signal at lag -2)

(2) **Reactive Signal (Return → Sentiment):** Conversely, Table 6.3 shows that for PLTR and BTC-USD, causality flows in the opposite direction. For PLTR, past returns "Granger-cause" future sentiment at lags 3, 4, and 5 ($p < 0.05$). For BTC-USD, this effect is more immediate, occurring at lags 1 and 2 ($p < 0.05$).

This indicates that, for these assets, social sentiment appears to be a reactive indicator ("echo") to past price movements. The visual analysis of the CCF, however, reveals distinct nuances: Figure 6.4 (PLTR) shows a weak and fluctuating correlation pattern at positive *lags*. A negative correlation is observed at lag 1, modest positive correlations at lags 2 and 3 (with 3 being the highest positive peak, but still low), and correlations close to zero or negative at lags 4 and 5. Although there is no strong and consistent positive peak, the Granger test suggests that the combined information from returns at these *lags* (3-5) still contributes to predicting sentiment. For BTC-USD (Figure 6.5), the CCF shows a clear and stronger positive correlation peak at lag 1, followed by a weaker positive correlation at lag 2. This visual pattern aligns well with the Granger test results, which found significance at both *lags*, indicating a reaction of sentiment to returns from the last one to two days.

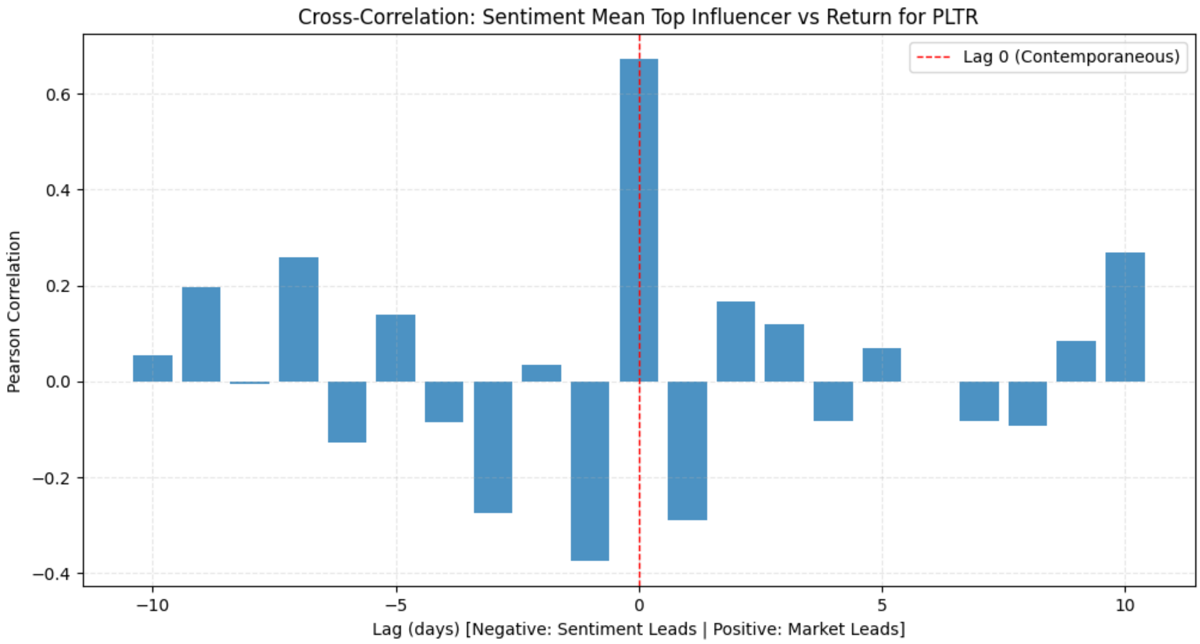

FIGURE 6.4. Cross-Correlation (PLTR): Sentiment Mean Top Influencer vs Return. (Reactive Signal at lags +1 to +5)

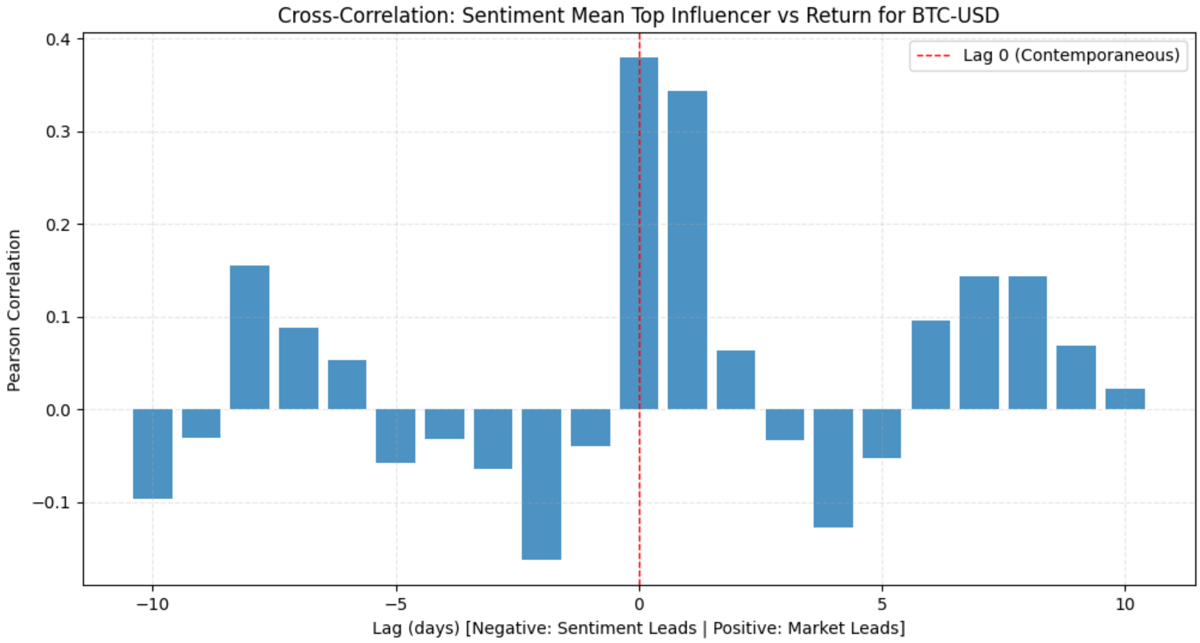

FIGURE 6.5. Cross-Correlation (BTC-USD): Sentiment Mean Top Influencer vs Return. (Reactive Signal at lags +1 and +2)

(3) **Absence of Causality:** For the remaining 13 assets (NIO Inc. (NIO), Lucid Group, Inc. (LCID), Intel Corporation (INTC), SOFI, etc.), no statistical evidence of Granger causality ($p > 0.05$) was found in either direction. This suggests that, for most of the analyzed symbols, the sentiment of "top influencers" and market returns move independently, or their relationship is purely contemporary (lag 0), not being captured by this causality test.

6.2. Regression Modeling: The Asymmetric Effect of Sentiment

To quantify the impact of social sentiment on asset returns, two regression models were applied: Linear Regression with Ordinary Least Squares (OLS) estimation and Quantile Regression (QR).

The OLS estimator provides an estimate of the average effect of sentiment, while quantile regression allows for a granular analysis of the impact at different points of the return distribution (e.g., on days of large drops vs. days of great euphoria).

The central hypothesis is that the OLS estimator, by focusing only on the mean, is insufficient to capture the complex dynamics of sentiment, and that the impact of sentiment is, in fact, asymmetric and non-linear.

6.2.1. Linear Regression: The Analysis of the Average Impact

The OLS estimator is expressed as:

$$Return_i = \beta_0 + \beta_1 \cdot Sentiment_i + \epsilon_i$$

Where $Return_i$ represents the asset's daily return, $Sentiment_i$ is the aggregated sentiment metric, β_1 is the coefficient of interest that measures the *average* impact of sentiment on returns, and ϵ_i is the error term.

TABLE 6.4. Results of OLS Estimation - Top Influencer Sentiment vs. Returns

Symbol	**Coefficient (β_1)**	**Std. Err.**	**p-value**	**R-squared**
INTC	0.170	0.032	0.000	0.310
NIO	0.163	0.026	0.000	0.369
PLTR	0.146	0.029	0.000	0.451
SOFI	0.123	0.028	0.000	0.273
LCID	0.123	0.021	0.000	0.345
XRP-USD	0.111	0.021	0.000	0.270
AMD	0.085	0.024	0.001	0.159
META	0.082	0.015	0.000	0.265
NKE	0.077	0.018	0.000	0.243
NVDA	0.073	0.019	0.000	0.289
GOOGL	0.068	0.012	0.000	0.313
UBER	0.066	0.012	0.000	0.256
AMZN	0.062	0.016	0.000	0.169
PFE	0.059	0.013	0.000	0.217
MSFT	0.058	0.012	0.000	0.213
DIA	0.046	0.014	0.002	0.180
BTC-USD	0.035	0.011	0.002	0.144

The results of the Linear Regression via OLS (Table 6.4) consistently indicate a positive and statistically significant average impact of "top influencers" sentiment on daily returns for all analyzed assets ($p < 0.01$). The explanatory power (R^2), however, varies substantially, from 14.4% for BTC-USD to an expressive 45.1% for PLTR. These average results confirm the general relevance of social sentiment but do not reveal the underlying dynamics under different market conditions.

The coefficient of determination (R^2) varies among the assets, indicating that the sentiment of top influencers explains a considerable portion of the variance in returns, being most prominent for PLTR ($R^2 = 0.451$) and NIO ($R^2 = 0.369$).

6.2.2. Quantile Regression: Revealing the Heterogeneity of Impact

Quantile regression, introduced by Koenker and Bassett (1978), allows for the estimation of the effect of sentiment at different quantiles (τ) of the return distribution. The model for the τ-th quantile is:

$$Q_\tau(Return_i \mid Sentiment_i) = \beta_{0\tau} + \beta_{1\tau} \cdot Sentiment_i$$

Where $\beta_{1\tau}$ measures the impact of sentiment on the τ-th quantile of returns. The analysis of the coefficients $\beta_{1\tau}$ for $\tau \in \{0.1, 0.25, 0.5, 0.75, 0.9\}$ reveals three distinct and economically significant patterns among the assets:

- **Decreasing Impact ("Buy the Dip")**: For a group of 5 assets (NVDA, Amazon.com, Inc. (AMZN), SOFI, Pfizer Inc. (PFE), NIKE, Inc. (NKE)), the impact of positive sentiment is significantly stronger in the lower quantiles (days of decline) and progressively decreases in the upper quantiles. As illustrated in Figure 6.6 for SOFI, the coefficient $\beta_{1\tau}$ at the 0.1 quantile is substantially larger than the OLS coefficient (mean) and the coefficients at the 0.75 or 0.9 quantiles. This suggests that, for these assets, the "bullish" sentiment of influencers functions more as a panic containment mechanism or a signal to "buy the dip," having less influence on days of euphoria.
- **Increasing Impact ("Fear Of Missing Out (FOMO)/Momentum")**: For another group of 4 assets (NIO, XRP-USD, PLTR, Uber Technologies, Inc. (UBER)), the opposite pattern is observed: the impact of positive sentiment is weaker in the lower quantiles and stronger in the upper quantiles (days of rise). Figure 6.7 for UBER exemplifies this dynamic, with the coefficient $\beta_{1\tau}$ increasing monotonically with τ, surpassing the OLS coefficient in the higher quantiles. This behavior suggests that sentiment acts as a momentum accelerator, amplifying euphoria and "FOMO" in already rising markets, a pattern often associated with more speculative assets.
- **Constant Impact ("Linear")**: Finally, for the largest group of 8 assets (META, BTC-USD, GOOGL, Microsoft Corporation (MSFT), INTC, Advanced Micro Devices, Inc. (AMD), LCID, SPDR Dow Jones Industrial Average ETF Trust (DIA)), the impact of sentiment is statistically stable across the different quantiles. As demonstrated in Figure 6.8 for META, the coefficients $\beta_{1\tau}$ (blue line) fluctuate around the OLS coefficient (red dashed line) and mostly remain within the OLS confidence interval (red shaded area). For these assets, including several Shares of large, financially stable companies with consistent performance (blue chips) and, interestingly, Bitcoin, the OLS estimator provides a reasonable and robust approximation of the sentiment-return relationship. Positive sentiment is consistently associated with positive returns, regardless of the market regime.

A fourth scenario that could also be identified is when the QR line has a "U" or inverted "U" shape. This pattern shows a complex non-linear relationship, for example, that sentiment only matters on days of extreme volatility (in the tails) and has no impact on normal days. However, none of the results exhibit a "U" or inverted "U" shape that is simultaneously pronounced and statistically significant (i.e., that clearly leaves the OLS confidence band in a non-monotonic way).

This quantile analysis unequivocally demonstrates that the relationship between social sentiment and market returns is heterogeneous and depends on the asset and market conditions (quantile). The mean estimated by OLS, although significant, masks these asymmetric dynamics.

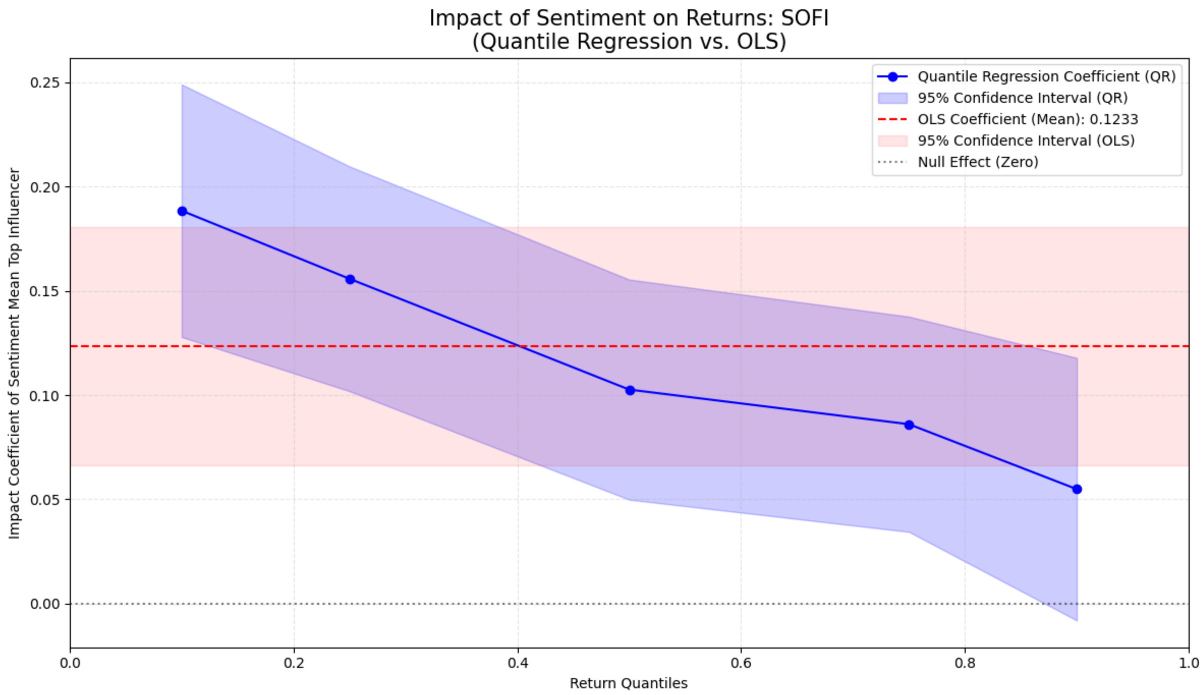

FIGURE 6.6. "Buy the Dip" Pattern: QR vs OLS Coefficients for SOFI. Decreasing impact with quantiles.

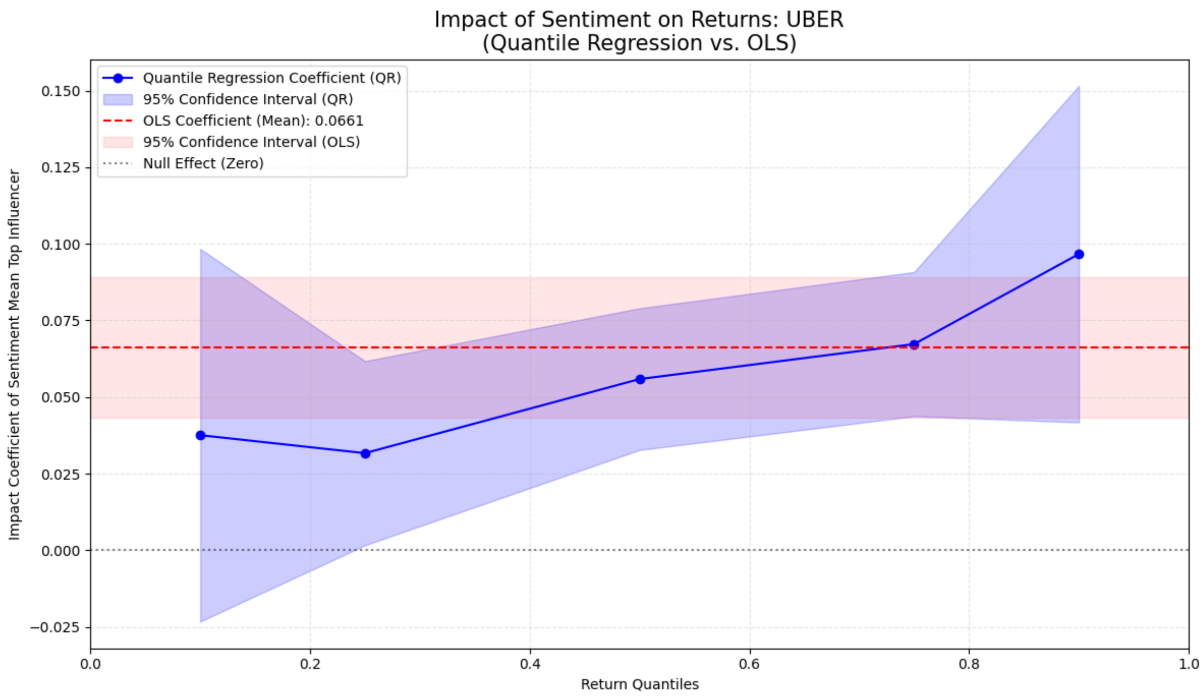

FIGURE 6.7. "FOMO/Momentum" Pattern: QR vs OLS Coefficients for UBER. Increasing impact with quantiles.

6.2.3. Economic Interpretation of the Results

The presented results have several important economic implications:

First, the consistent statistical significance of the sentiment coefficient (β_1 in OLS and $\beta_{1\tau}$ in QR) reinforces the evidence that social sentiment—particularly that from

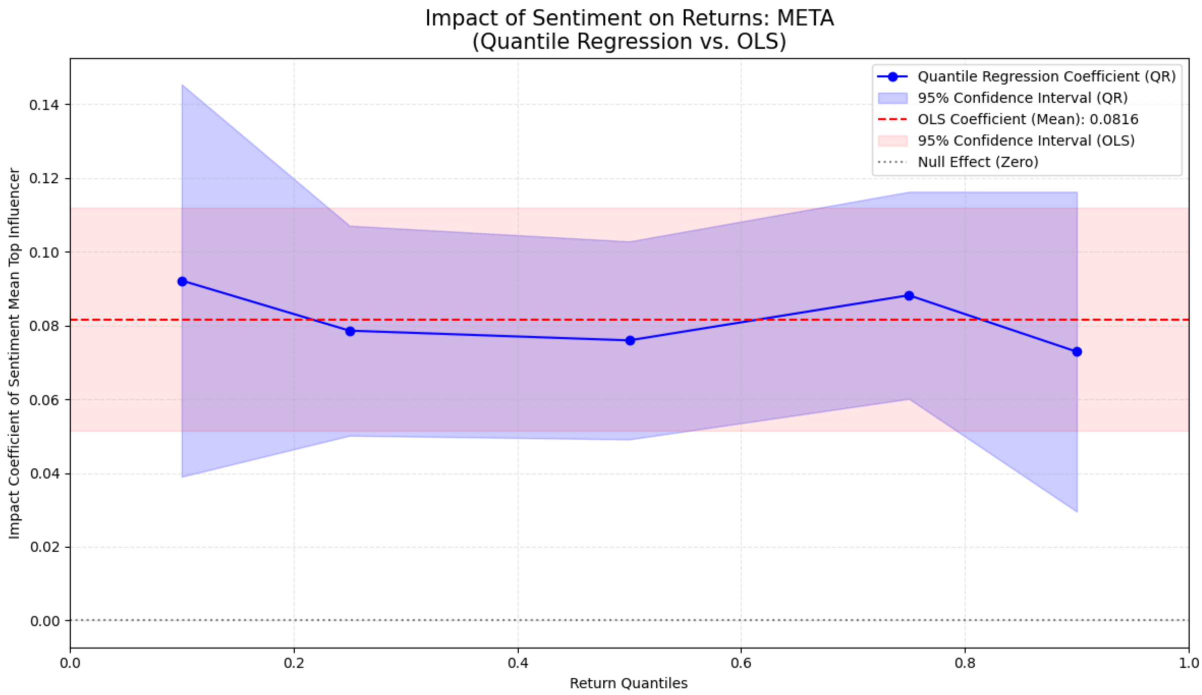

FIGURE 6.8. "Constant-Linear" Pattern: QR vs OLS Coefficients for META. Stable impact across quantiles.

influencers—contains economically relevant information for financial markets. This result is in line with the *Adaptive Markets Hypothesis* (Lo, 2004), according to which behavioral factors and rationality mechanisms coexist and evolve according to the informational context.

In this context, social media acts as dynamic channels for disseminating this information (or correlated noise), reflecting the continuous adaptation of market participants to new data flows and collective perceptions.

Thus, the finding is also compatible with the *Informational Efficiency Hypothesis* (Fama, 1970), as information originating in the social environment—when relevant—tends to be quickly incorporated into asset prices, contributing to the adaptive efficiency of markets.

Second, the asymmetry observed in the quantile regression indicates that social sentiment can contribute to the formation of speculative bubbles. The "Buy the Dip" pattern suggests that, for certain assets (e.g., NVDA, SOFI), sentiment may have a stabilizing role during declines, perhaps reflecting fundamental confidence or opportunism. In contrast, the "FOMO/Momentum" pattern (e.g., NIO, Ripple (XRP), PLTR) points to a destabilizing role of sentiment, amplifying trends and potentially contributing to the formation of speculative bubbles or excessive euphoria. The fact that this pattern includes "meme stocks" and alternative cryptocurrencies (XRP) is particularly suggestive.

Third, the robustness of the "Constant-Linear" pattern for blue chips (META, GOOGL, MSFT) and Bitcoin suggests that, for assets with higher liquidity, more widespread information, and perhaps a more diversified investor base, the sentiment-return relationship is more stable and less susceptible to extreme dynamics dependent on the market regime. The behavior of Bitcoin in this group, distinct from XRP, is notable.

6.2.4. Limitations and Robustness of the Results

The limitations inherent to studies based on social media sentiment apply: the representativeness of the Stocktwits sample, as it is based on a specific sample of platform users, which may not be representative of the broader universe of investors; the accuracy of sentiment classification, which may contain errors that attenuate the observed correlations; and the possibility of omitted variables. However, the QR analysis, by itself, increases robustness by not relying on assumptions of error normality, which are often violated in financial data. The consistency of the patterns found across multiple assets within each group ("Buy the Dip," "FOMO," "Constant") also lends greater confidence to the conclusions. The confirmation of the adequacy of OLS by QR for the "Constant-Linear" group also serves as an internal robustness test.

6.2.5. Implications for Investors and Regulators

The findings of this study have important practical implications for different **stakeholders** in the financial market. For investors, the results suggest that social sentiment can be incorporated into investment strategies as a complementary indicator. The observed predictive capacity, particularly for cryptocurrencies, can be explored through market *timing* strategies.

For regulators, the results raise questions about the need to monitor social media as part of financial stability oversight. The disproportionate influence of "top influencers" suggests that specific regulation for financial communications on social media may be necessary, especially when they can significantly affect asset prices.

CHAPTER 7

Conclusions

7.1. Main Findings

The comprehensive analysis of the data reveals several important discoveries about the relationship between social sentiment and financial market behavior. First, the existence of significant correlations between sentiment metrics extracted from the Stocktwits platform and the returns of various financial assets is confirmed, with cryptocurrencies exhibiting the highest correlations.

Second, Granger causality tests demonstrate that social sentiment, particularly that of top influencers, has predictive power over the future returns of certain assets. This predictive capacity is more pronounced for assets with greater retail investor participation and lower informational efficiency.

Third, quantile regression analysis reveals important asymmetries in the impact of sentiment, with differentiated effects at different points of the return distribution. This heterogeneity suggests that social sentiment can contribute to both the formation of speculative bubbles and market corrections.

7.2. Theoretical Implications

The results obtained have important implications for several established financial theories. The evidence of social sentiment's predictive power challenges the strong form of the efficient market hypothesis, suggesting that publicly available information on social media is not immediately incorporated into asset prices.

Simultaneously, the findings are consistent with behavioral finance theories that emphasize the role of emotions and cognitive biases in investment decision-making. The manifestation of these biases on social media and their subsequent influence on asset prices provides additional empirical evidence for these theories.

The observed heterogeneity among different asset classes also supports theories that emphasize the importance of market structure and investor composition in determining informational efficiency. Markets with greater institutional participation show less sensitivity to social sentiment, while markets dominated by retail investors exhibit greater influence.

7.3. Practical Implications

For investors, the results suggest that social sentiment can be incorporated as a complementary factor in investment strategies, particularly for assets with high sensitivity to this factor. However, the temporal variability in predictive capacity requires a careful and adaptive approach.

For asset managers, monitoring social sentiment can provide valuable insights into investor expectations and perceptions, allowing for better risk management and identification of investment opportunities. The ability to anticipate market movements based on sentiment can be particularly valuable in volatile markets.

For regulators, the findings raise important questions about the need to monitor social media as part of financial stability oversight. The disproportionate influence of certain influential users suggests that specific regulation for financial communications on social media may be necessary.

7.4. Study Limitations

Despite the significant results, it is important to recognize several limitations of this study. First, the analysis is based on a specific sample of Stocktwits platform users, which may not be representative of the broader universe of investors. Generalization of the results to other platforms or investor populations requires caution.

Second, the automatic sentiment classification, although based on advanced natural language processing techniques, may contain errors that attenuate the observed correlations. Manual validation of a subsample of the classifications suggests high accuracy, but systematic errors cannot be completely ruled out.

Third, the analysis period, although comprehensive, may not capture all relevant market conditions. Periods of financial crisis or extreme events may reveal different patterns in the relationship between sentiment and returns.

7.5. Directions for Future Research

The results of this study open several promising avenues for future research. First, extending the analysis to other social media platforms, such as Twitter, Reddit, or LinkedIn, could provide a more comprehensive perspective on investor sentiment.

Second, the incorporation of more advanced natural language processing techniques, including large language models, could improve the accuracy of sentiment classification and capture more subtle nuances in investor communications.

Third, social network analysis could reveal patterns of influence and sentiment propagation that are not captured by the aggregated metrics used in this study. The identification of key users and influence paths could provide valuable insights into the mechanisms of sentiment transmission.

Fourth, extending the analysis to international markets and different time zones could reveal global patterns in the relationship between social sentiment and market behavior. The synchronization or desynchronization of sentiment across different geographical regions constitutes a particularly interesting area of investigation.

7.6. Final Considerations

This study presented a comprehensive analysis of the relationship between social sentiment extracted from the Stocktwits platform and financial market behavior. Through a combination of advanced statistical techniques and detailed visualizations, it was possible to identify significant patterns and quantify the impact of social sentiment on asset returns.

The results confirm the existence of causal relationships between social sentiment and market returns, with important variations among different asset classes. Cryptocurrencies emerge as the class most sensitive to social sentiment, followed by technology stocks and, finally, traditional market indices.

The quantile regression analysis reveals important asymmetries in the impact of sentiment, suggesting that it can contribute to both the formation of speculative bubbles and market corrections. This heterogeneity has important implications for risk management and investment strategies.

The findings have significant implications for investors, asset managers, and regulators, providing empirical evidence on the growing role of social media in modern financial markets. Simultaneously, the identified limitations and directions for future research highlight the need for additional studies to deepen the understanding of these complex phenomena.

In summary, this study contributes to the growing body of literature on the impact of social media on financial markets, providing robust empirical evidence on the relevance of social sentiment as an explanatory factor for market movements. The results suggest that, although financial markets maintain a significant degree of efficiency, the influence of collective emotions and perceptions, manifested through social media, constitutes an important factor that cannot be ignored in the analysis and forecasting of market behavior.

Bibliography

Akkerman, S. (2023). Fintwitbert-sentiment [Obtido de Hugging Face]. https://huggingface.co/StephanAkkerman/FinTwitBERT-sentiment

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.

Araci, D. (2019). Finbert: Financial sentiment analysis with pre-trained language models. *arXiv preprint arXiv:1908.10065*. https://arxiv.org/abs/1908.10065

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

Bakhtiar, T., Luo, X., & Adelopo, I. (2023). The impact of fundamental factors and sentiments on the valuation of cryptocurrencies. *Blockchain: Research and Applications*, *4*(2), 100154. https://doi.org/10.1016/j.bcra.2023.100154

Bakshy, E., Rosenn, I., Marlow, C., & Adamic, L. A. (2011). Everyone's an influencer: Quantifying influence on twitter. *Proceedings of the fourth ACM international conference on Web search and data mining*, 65–74. https://doi.org/10.1145/1935826.1935845

Bird, R., Gallagher, D. R., Khan, A., & Yeung, D. (2023). Do emotions influence investor behavior? *Journal of Behavioral Finance*, 1–22. https://doi.org/10.1080/15427560.2023.2282966

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

Campbell, J. Y., Lo, A. W., & MacKinlay, A. C. (1997). *The econometrics of financial markets*. Princeton University Press.

CardiffNLP. (2023). Twitter roberta base sentiment latest [computer software] [Hugging Face repository. Accessed: 2025-10-21].

CERN & OpenAIRE. (n.d.). Zenodo — open access research data repository [Accessed: 2025-10-28].

Chapman, P., Clinton, J., Kerber, R., Khabaza, T., Reinartz, T., Shearer, C., & Wirth, R. (2000). *Crisp-dm 1.0: Step-by-step data mining guide* (tech. rep.). The CRISP-DM Consortium. https://www.the-modeling-agency.com/crisp-dm.pdf

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

Chernozhukov, V., & Hansen, C. (2005). An iv model of quantile treatment effects. *Econometrica*, *73*(1), 245–261. https://doi.org/10.1111/j.1468-0262.2005.00594.x

Cioroianu, I., Conlon, T., Corbet, S., Larkin, C., Oxley, L., & Taffler, R. (2024). The influence of social media emotions: Evidence from the kodak manic episode. *Journal of Behavioral Finance*. https://doi.org/10.1080/15427560.2024.2385902

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

Engle, R. F. (1982). Autoregressive conditional heteroscedasticity with estimates of the variance of united kingdom inflation. *Econometrica*, *50*(4), 987–1007. https://doi.org/10.2307/1912773

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

Fan, J. H., Binnewies, S., & De Silva, S. (2023). Wisdom of crowds and commodity pricing. *Journal of Futures Markets*, *43*(8), 1040–1068. https://doi.org/10.1002/fut.22393

FiniteAutomata. (2021). Bertweet base sentiment analysis [computer software] [Hugging Face repository. Accessed: 2025-10-08].

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

Granger, C. W. J. (1980). Testing for causality: A personal viewpoint. *Journal of Economic Dynamics and Control*, *2*(1), 329–352. https://doi.org/10.1016/0165-1889(80)90069-X

Hamilton, J. D. (1994). *Time series analysis*. Princeton University Press.

Koenker, R. (2005). *Quantile regression*. Cambridge University Press. https://doi.org/10.1017/CBO9780511754098

Koenker, R., & Bassett, G. (1978). Regression quantiles. *Econometrica*, *46*(1), 33–50. https://doi.org/10.2307/1913643

Kristoufek, L. (2013). Bitcoin meets google trends and wikipedia: Quantifying the relationship between phenomena of the internet era. *Scientific Reports*, *3*, 3415. https://doi.org/10.1038/srep03415

Li, T., Chen, H., Liu, W., & Yu, G. (2023). Understanding the role of social media sentiment in identifying irrational herding behavior in the stock market. *International Review of Economics and Finance*, *87*, 163–179. https://doi.org/10.1016/j.iref.2023.04.016

Lo, A. W. (2004). The adaptive markets hypothesis: Market efficiency from an evolutionary perspective. *The Journal of Portfolio Management*, *30*(5), 15–29. https://doi.org/10.3905/jpm.2004.442611

Loughran, T., & McDonald, B. (2011). When is a word "negative"? *Financial Analysts Journal*, *67*(2), 43–55. https://doi.org/10.2469/faj.v67.n2.4

Page, M. J., McKenzie, J. E., Bossuyt, P. M., Boutron, I., Hoffmann, T. C., Mulrow, C. D., Shamseer, L., Tetzlaff, J. M., Akl, E. A., Brennan, S. E., Chou, R., Glanville, J., Grimshaw, J. M., Hróbjartsson, A., Lalu, M. M., Li, T., Loder, E. W., Mayo-Wilson, E., McDonald, S., … Moher, D. (2021). The prisma 2020 statement: An updated guideline for reporting systematic reviews. *BMJ*, *372*, n71. https://doi.org/10.1136/bmj.n71

Pearson, K. (1895). Note on regression and inheritance in the case of two parents. *Proceedings of the Royal Society of London*, *58*, 240–242. https://doi.org/10.1098/rspl.1895.0041

ProsusAI. (2020). Finbert: Financial sentiment analysis model [computer software] [Hugging Face repository. Accessed: 2025-10-21].

Sabherwal, R., Gupta, A., & Das, S. R. (2011). Information quality and social media: A study of an algorithmic trading. *Journal of Management Information Systems*, *28*(3), 19–54.

Santa Ana, W. (2025, October). *Sentiment analysis in financial social media: Impact on market volatility* (Version v1.0.0). Zenodo. https://doi.org/10.5281/zenodo.17468843

SCImago. (n.d.). Sjr — scimago journal & country rank [portal] [Accessed: 2024-11-05].

Scimago Journal & Country Rank. (n.d.). Scimago journal & country rank [Retrieved January 3, 2025].

Shiller, R. J. (2003). *Irrational Exuberance* (Second Edition or relevant edition for 2003, check original source if possible - First edition was 2000, Second 2005. Using 2003 often refers to reprints or specific versions.). Princeton University Press.

Spearman, C. (1904). The proof and measurement of association between two things. *The American Journal of Psychology*, *15*(1), 72–101. https://doi.org/10.2307/1412159

Stocktwits. (2025). *Stocktwits [portal]*. https://stocktwits.com/

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

Vaswani, A., Shazeer, N., Parmar, N., Uszkoreit, J., Jones, L., Gomez, A. N., Kaiser, Ł., & Polosukhin, I. (2017). Attention is all you need. *Advances in Neural Information Processing Systems*, *30*. https://papers.nips.cc/paper/7181-attention-is-all-you-need

Wooldridge, J. M. (2016). *Introductory econometrics: A modern approach* (6th ed.). Cengage Learning.

Wu, Y., Wu, S., Xu, F., & Jiang, J. (2024). Wisdom of crowds or awkward squad? social interaction and the information efficiency of the chinese capital market. *Research in International Business and Finance*, *71*, 102486. https://doi.org/10.1016/j.ribaf.2024.102486

Yadav, A., & Chakraborty, A. (2023). Synthesis of the effect of investor sentiment on stock returns: A systematic review. *Global Business and Economics Review*, *28*(4), 425–446. https://doi.org/10.1504/GBER.2023.132778

Yahoo Finance. (n.d.). Historical data [Accessed: 2025-06-17].

Yang, Z., Dai, Z., Yang, Y., Carbonell, J., Salakhutdinov, R. R., & Le, Q. V. (2020). Xlnet: Generalized autoregressive pretraining for language understanding. *Advances in Neural Information Processing Systems*, *32*. https://papers.nips.cc/paper/2019/hash/dc6a7e655d7e5840e66733e9ee67cc69-Abstract.html

APPENDIX A

Cross-Correlation

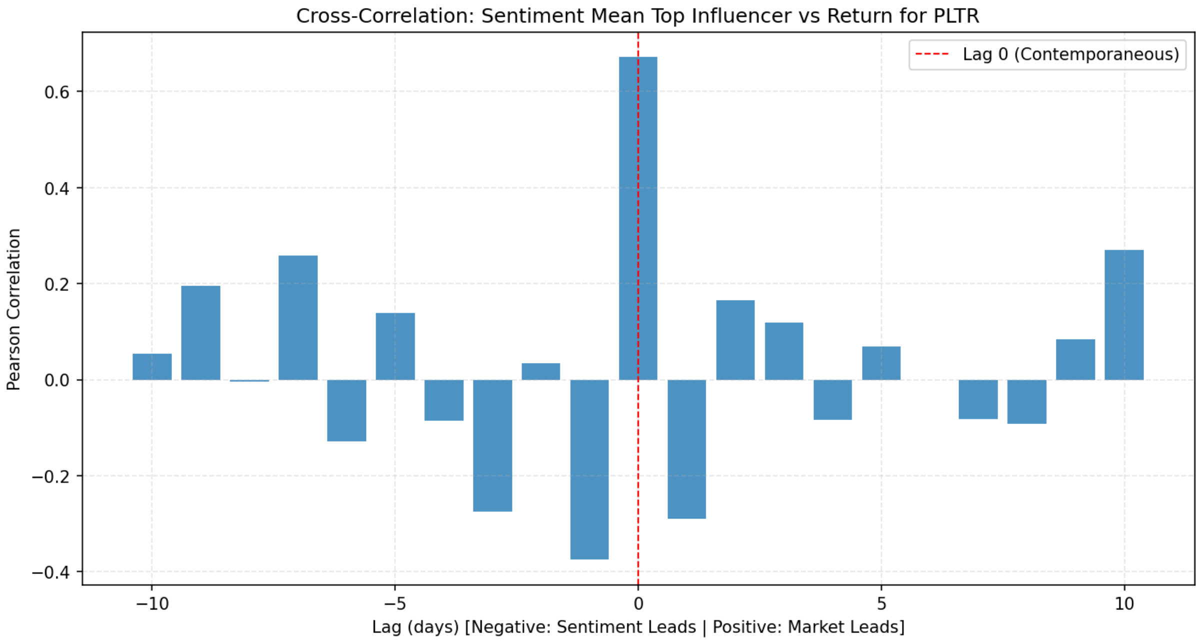

Figure A.1. Cross-Correlation: Sentiment Mean Top Influencer vs Return for PLTR

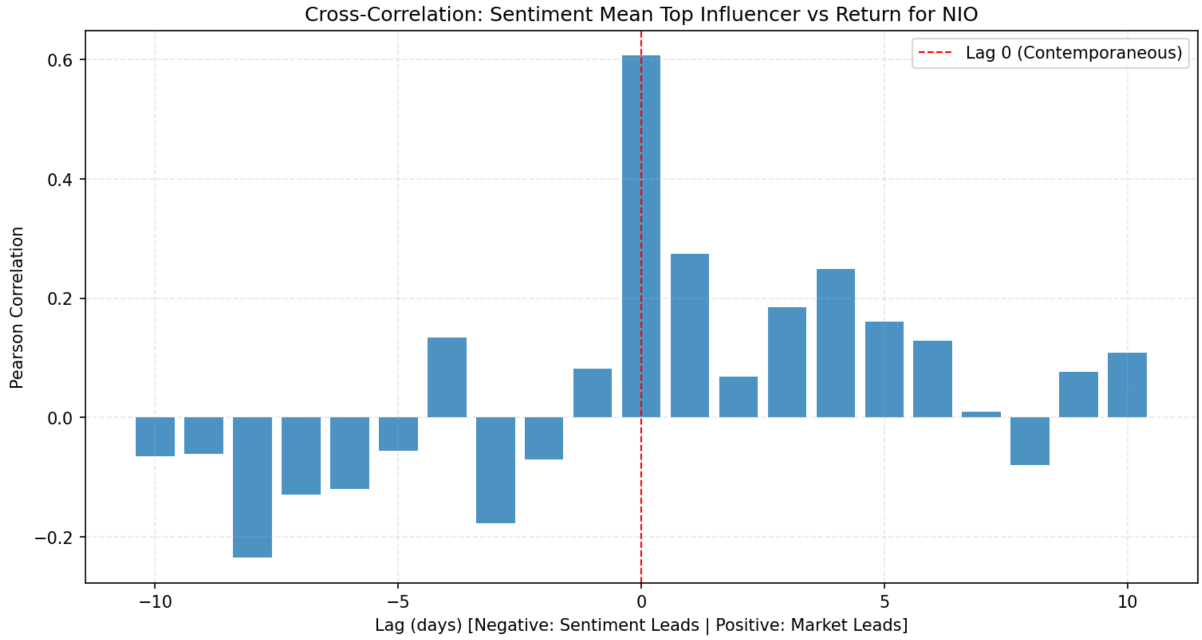

FIGURE A.2. Cross-Correlation: Sentiment Mean Top Influencer vs Return for NIO

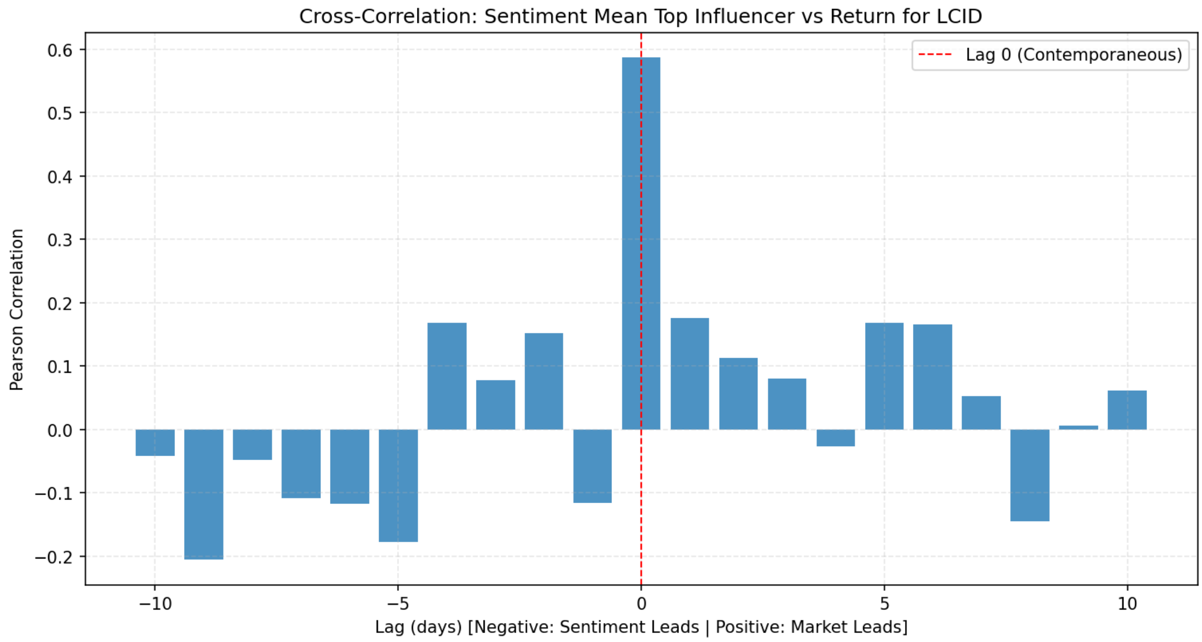

FIGURE A.3. Cross-Correlation: Sentiment Mean Top Influencer vs Return for LCID

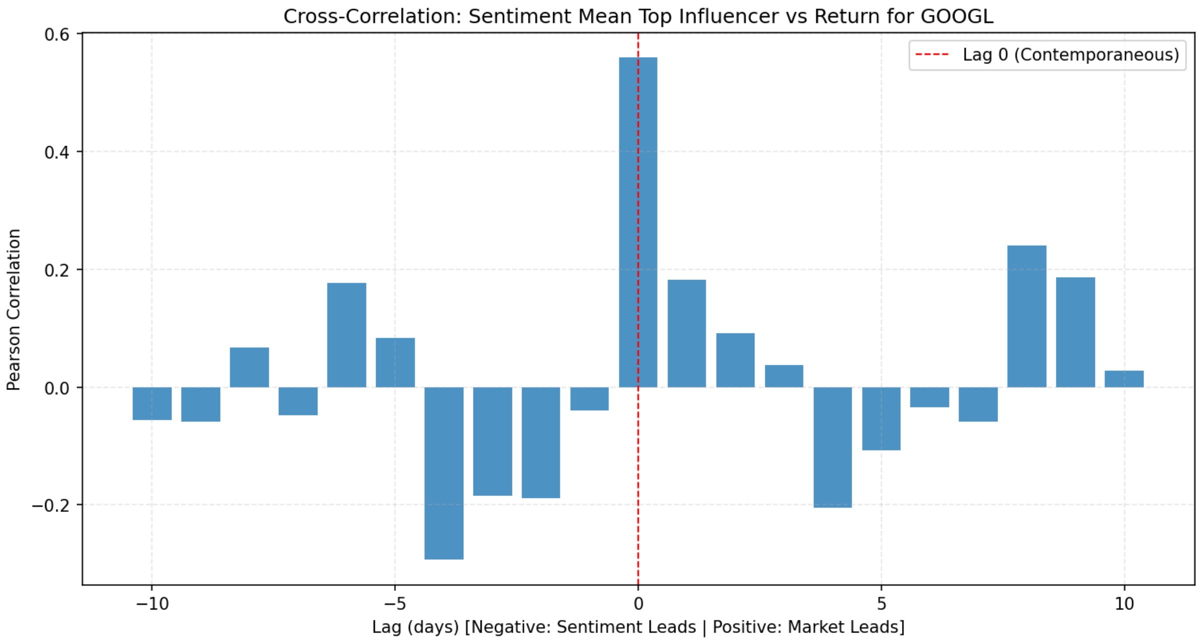

FIGURE A.4. Cross-Correlation: Sentiment Mean Top Influencer vs Return for GOOGL

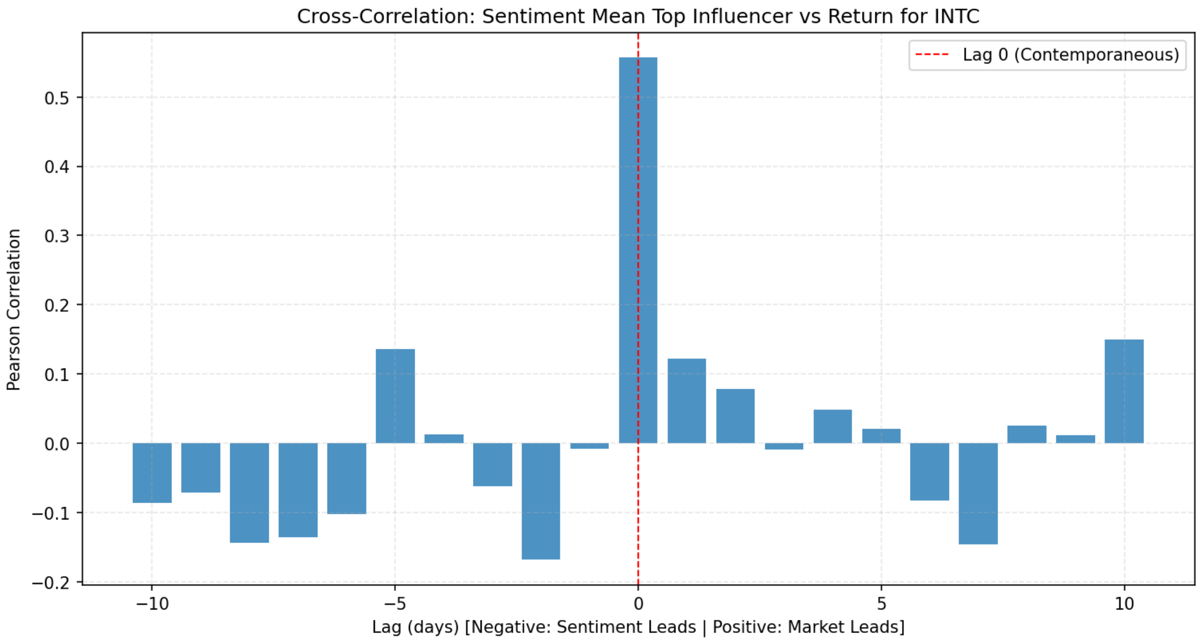

FIGURE A.5. Cross-Correlation: Sentiment Mean Top Influencer vs Return for INTC

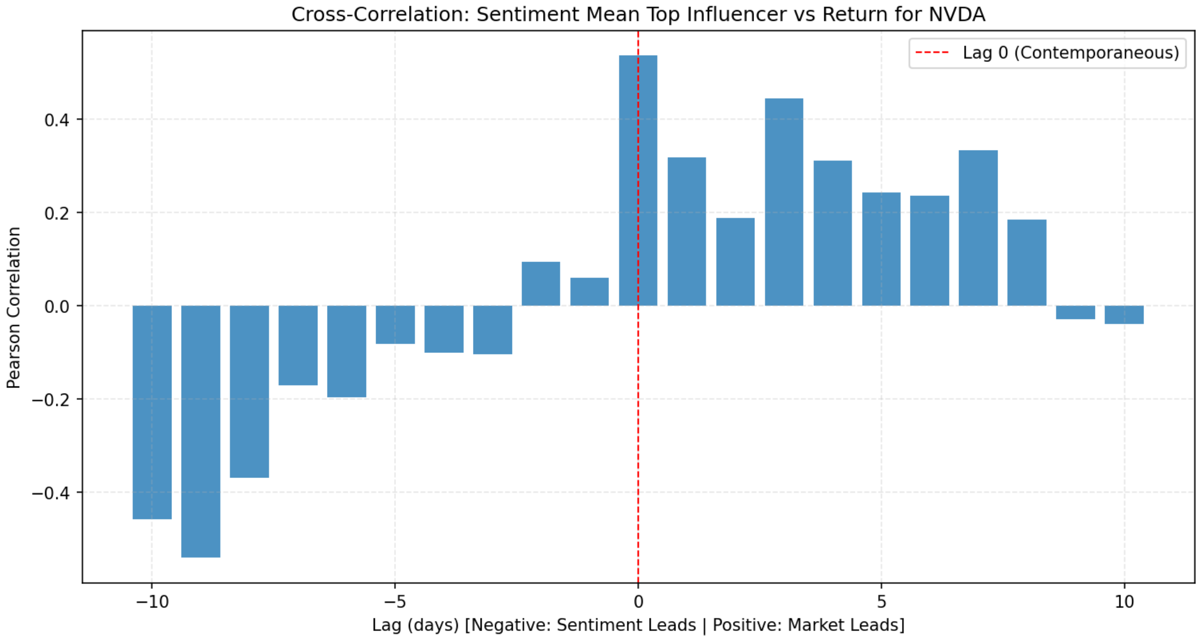

FIGURE A.6. Cross-Correlation: Sentiment Mean Top Influencer vs Return for NVDA

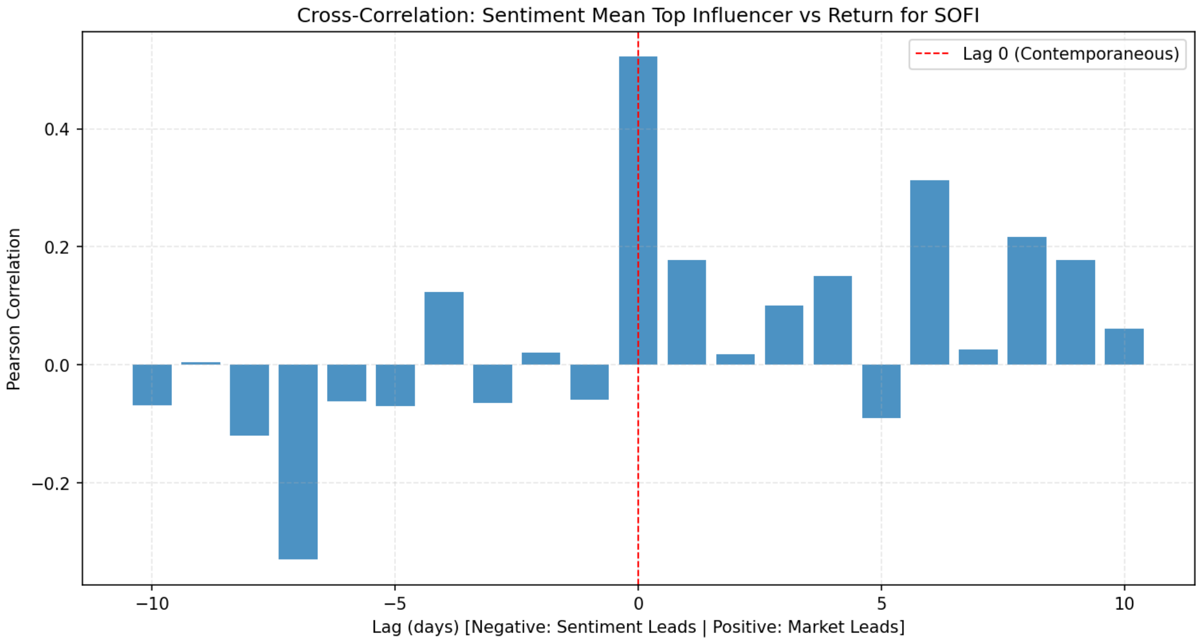

FIGURE A.7. Cross-Correlation: Sentiment Mean Top Influencer vs Return for SOFI

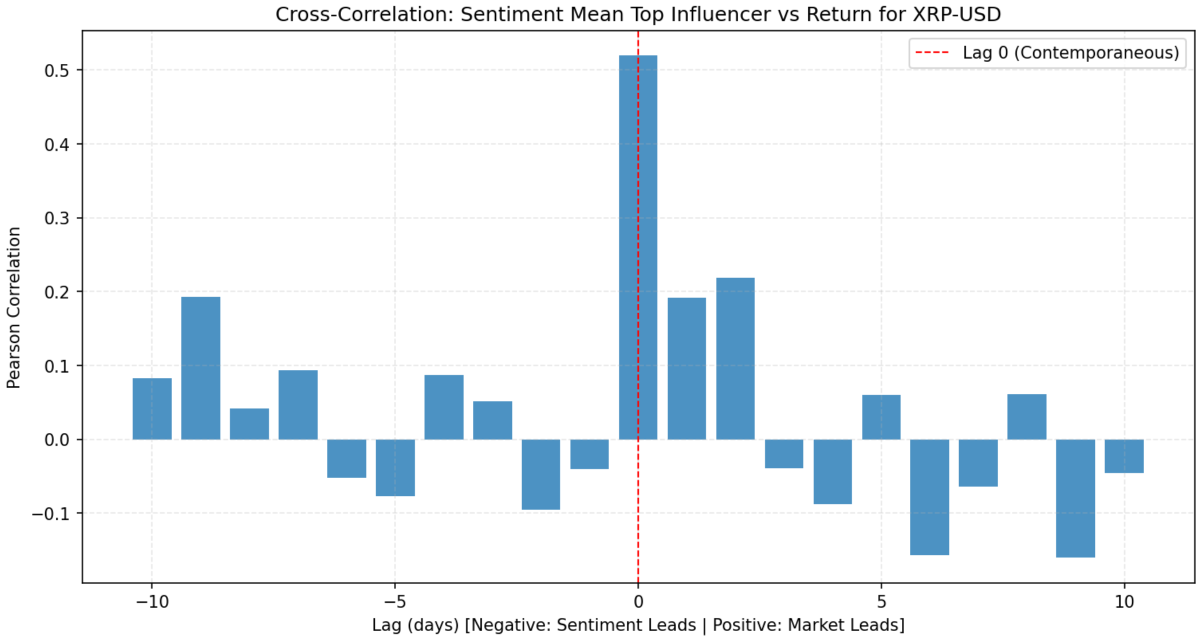

FIGURE A.8. Cross-Correlation: Sentiment Mean Top Influencer vs Return for XRP-USD

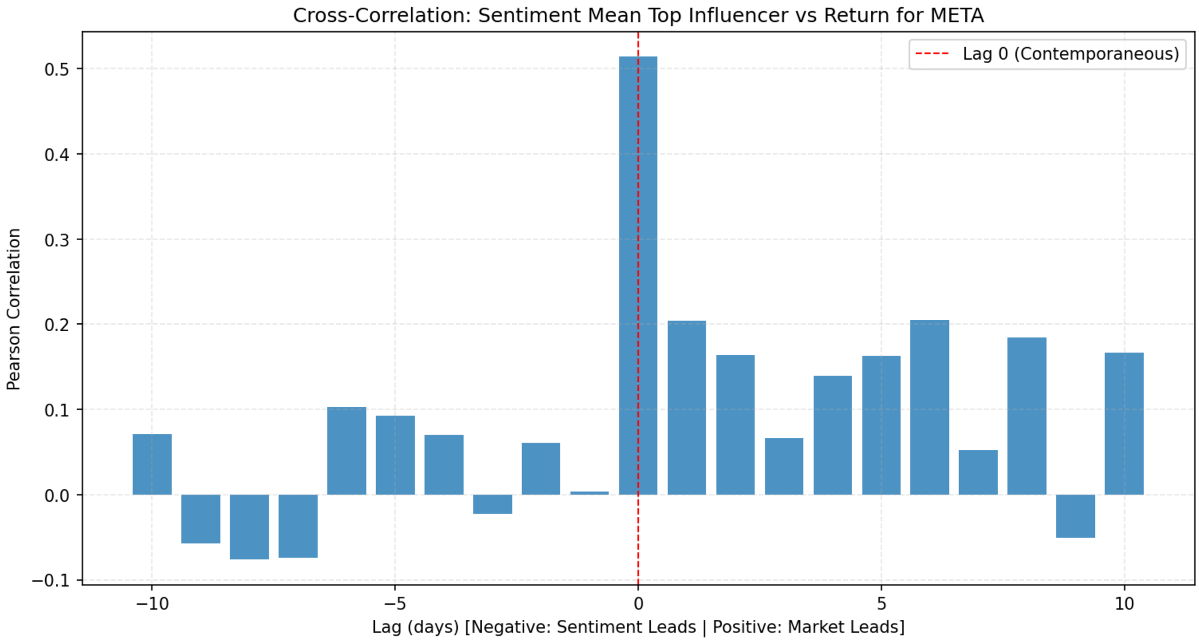

FIGURE A.9. Cross-Correlation: Sentiment Mean Top Influencer vs Return for META

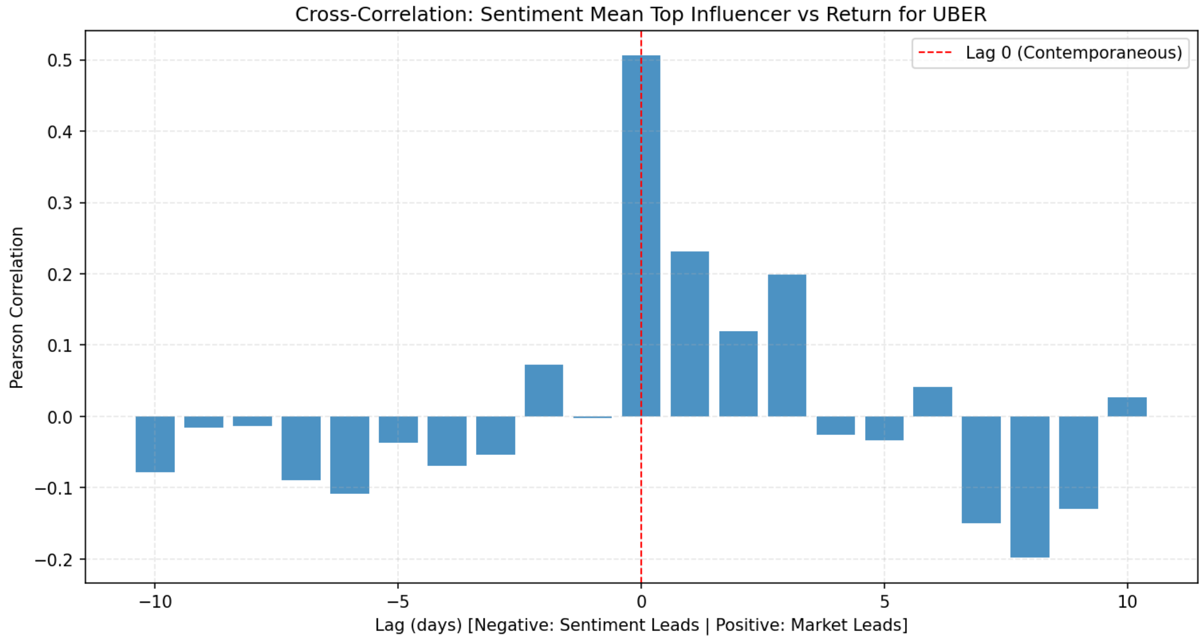

FIGURE A.10. Cross-Correlation: Sentiment Mean Top Influencer vs Return for UBER

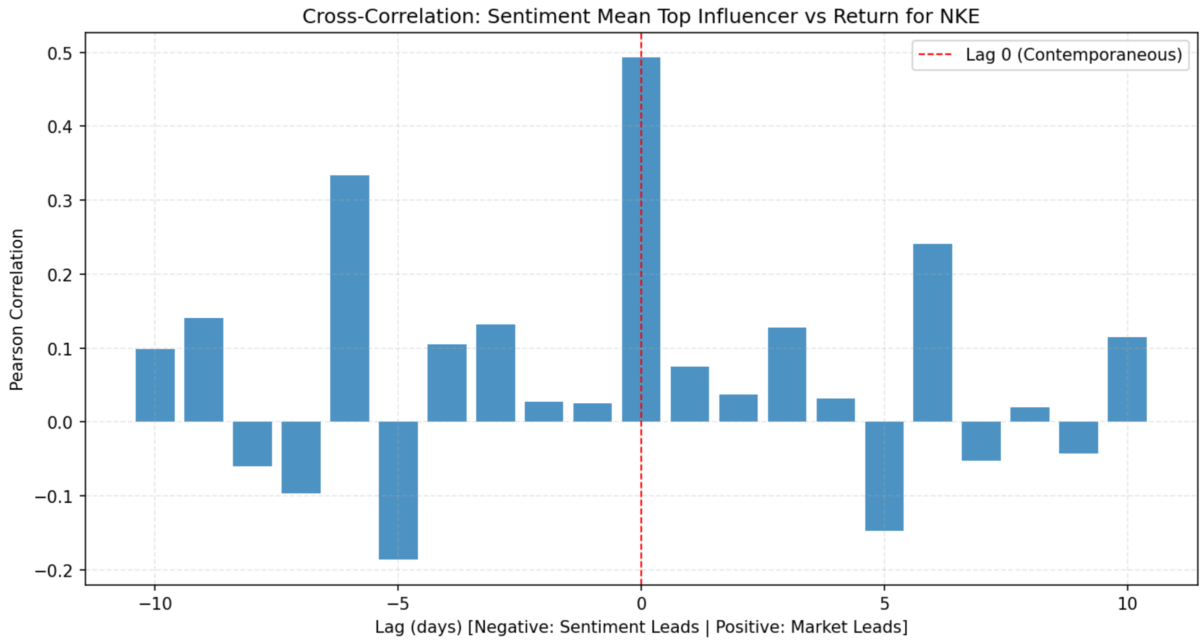

FIGURE A.11. Cross-Correlation: Sentiment Mean Top Influencer vs Return for NKE

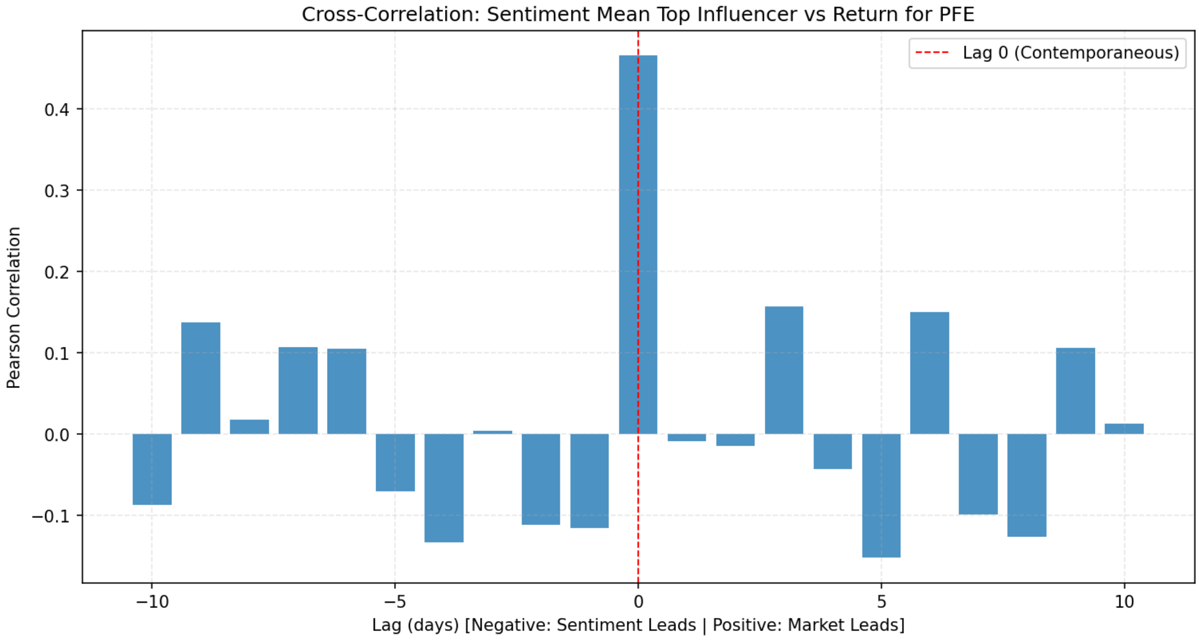

FIGURE A.12. Cross-Correlation: Sentiment Mean Top Influencer vs Return for PFE

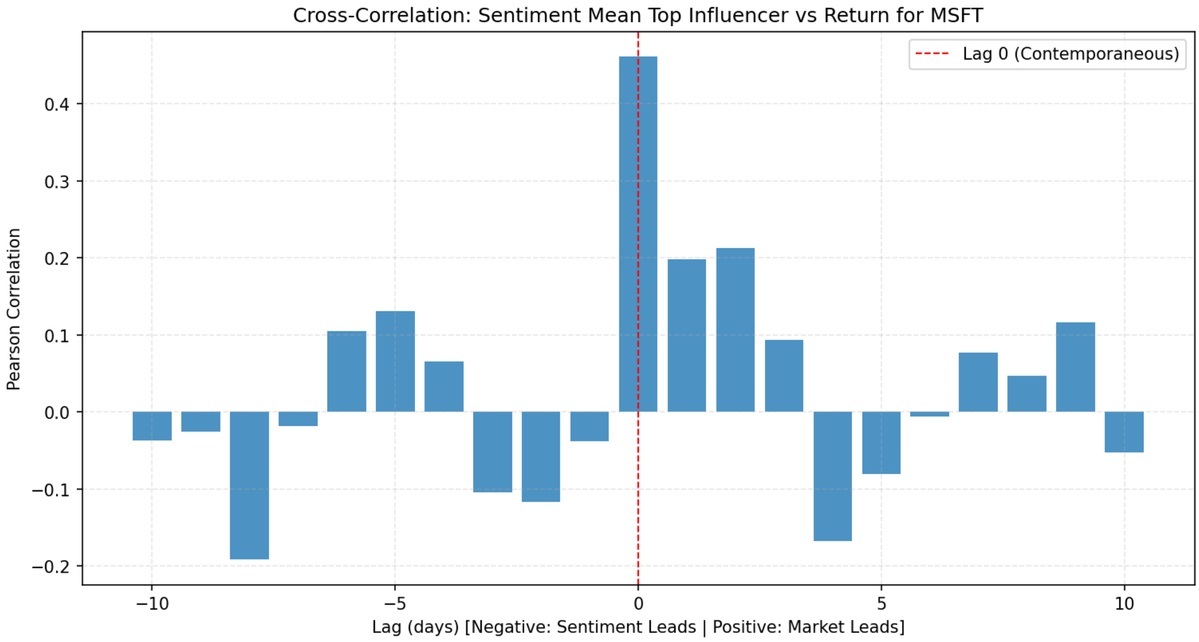

FIGURE A.13. Cross-Correlation: Sentiment Mean Top Influencer vs Return for MSFT

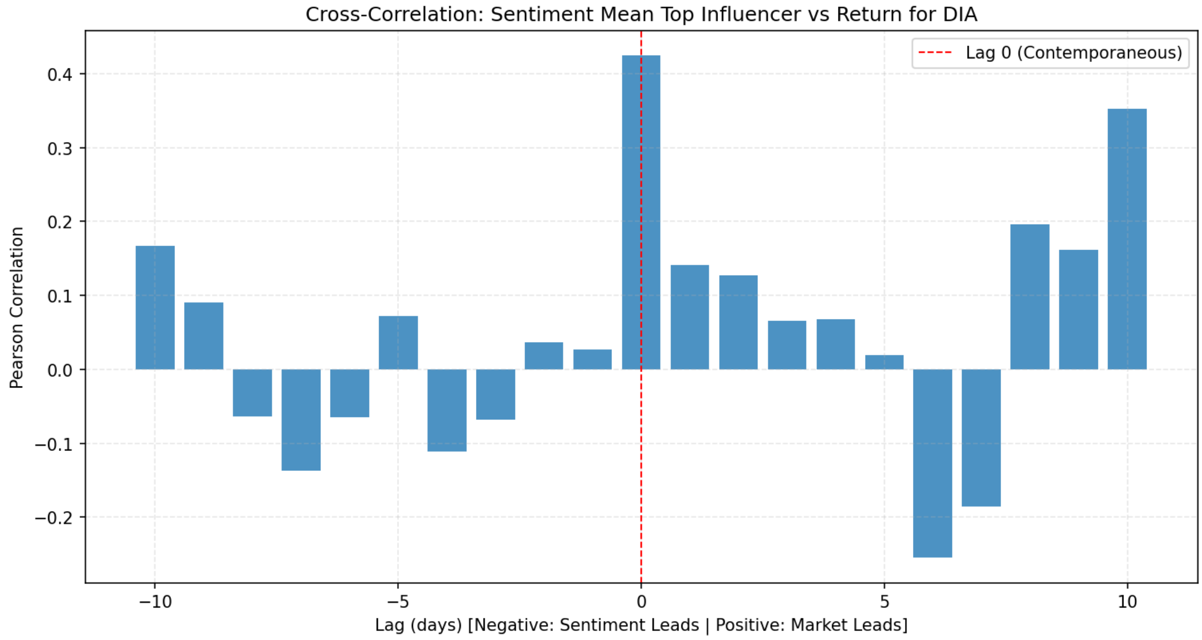

FIGURE A.14. Cross-Correlation: Sentiment Mean Top Influencer vs Return for DIA

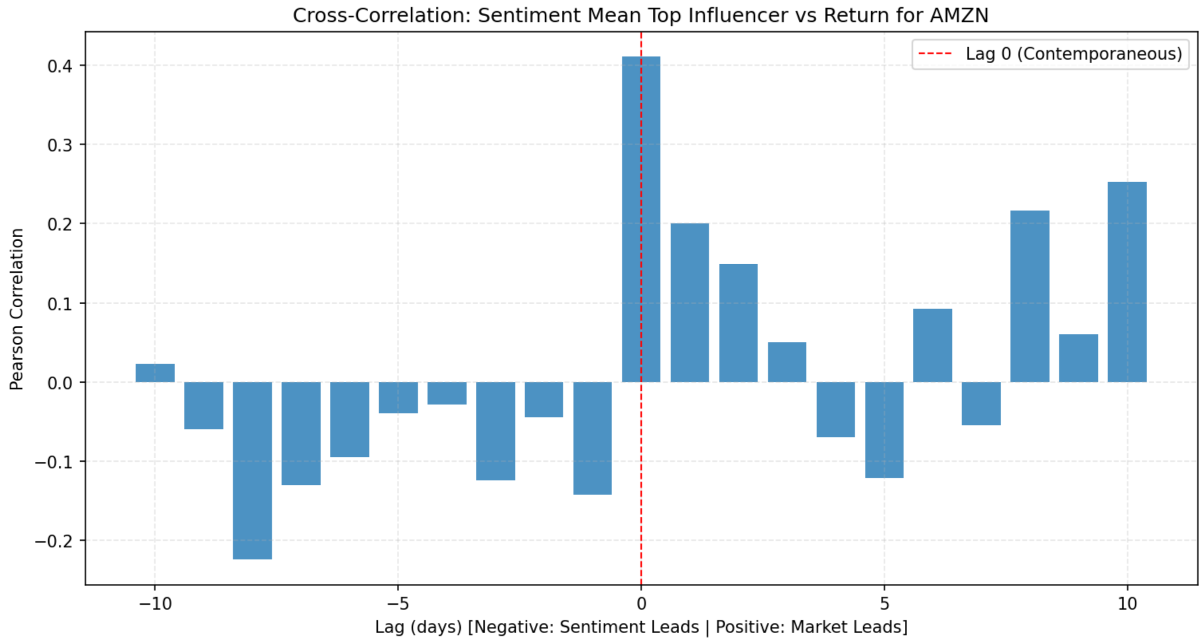

FIGURE A.15. Cross-Correlation: Sentiment Mean Top Influencer vs Return for AMZN

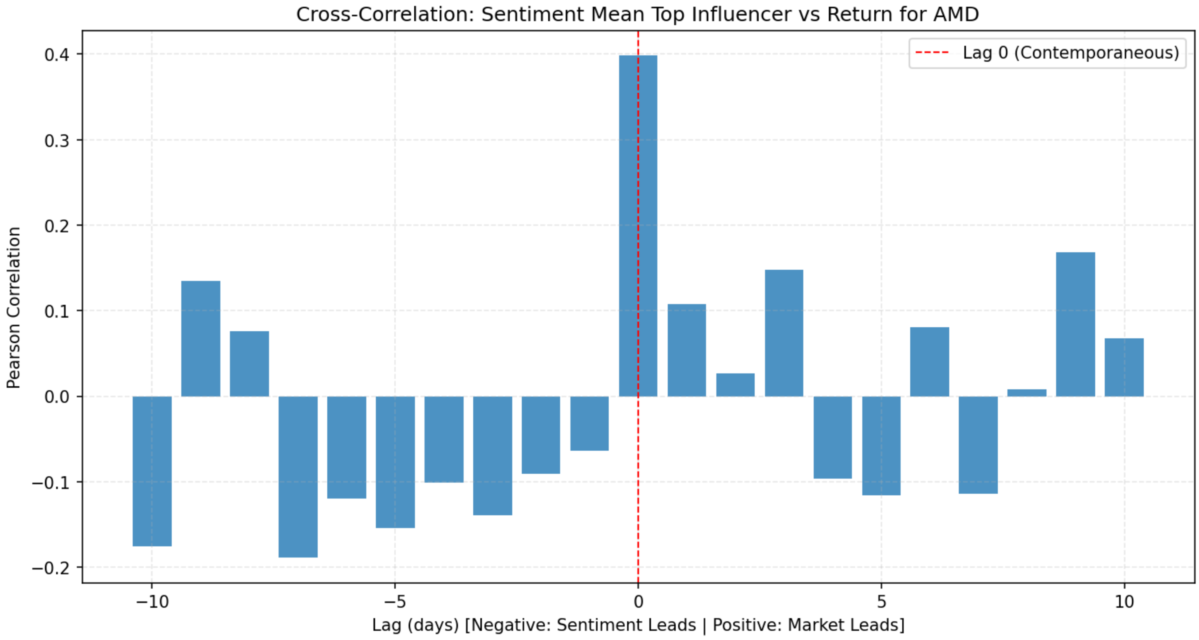

FIGURE A.16. Cross-Correlation: Sentiment Mean Top Influencer vs Return for AMD

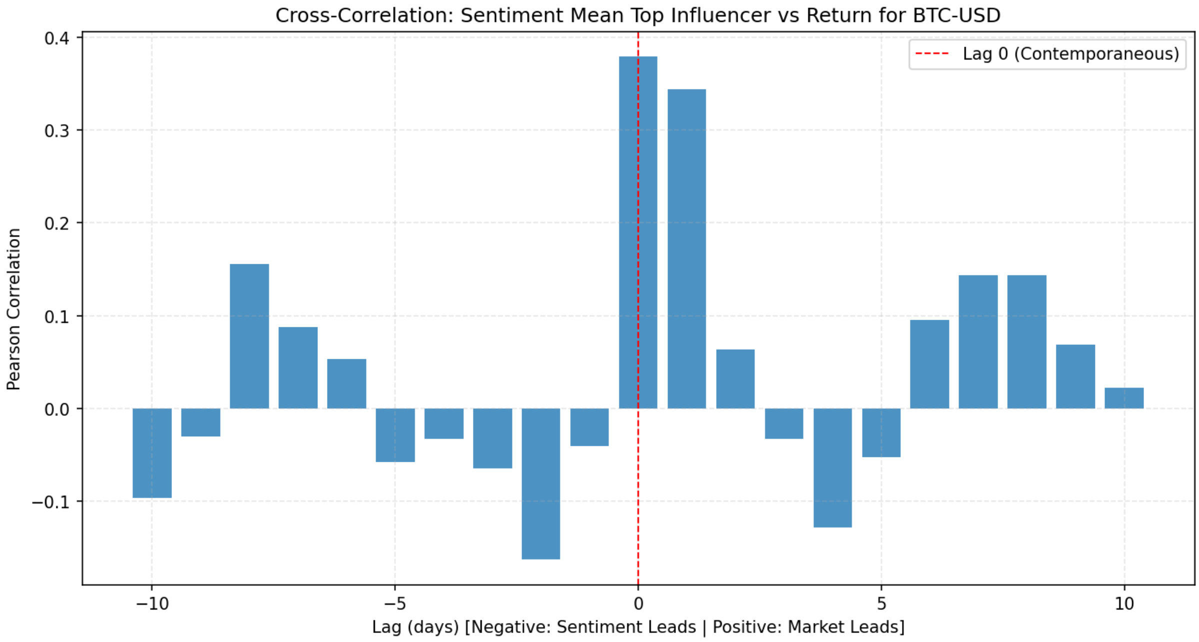

FIGURE A.17. Cross-Correlation: Sentiment Mean Top Influencer vs Return for BTC-USD

APPENDIX B

Quantile Regression and Linear Regression with OLS Estimator

Table B.1. Quantile 0.1 — NVDA

Dep. Variable:	Return	**Pseudo R-squared:**	0.2461
Model:	QuantReg	**Bandwidth:**	nan
Method:	Least Squares	**Sparsity:**	nan
Date:	Mon, 27 Oct 2025	**No. Observations:**	38
Time:	08:58:23	**Df Residuals:**	36
		Df Model:	1

	coef	**std err**	**t**	**P> |t|**	**[0.025**	**0.975]**
Intercept	-0.0450	nan	nan	nan	nan	nan
sentiment_mean_top_influencer	0.0976	nan	nan	nan	nan	nan

Table B.2. Quantile 0.25 — NVDA

Dep. Variable:	Return	**Pseudo R-squared:**	0.2162
Model:	QuantReg	**Bandwidth:**	0.04328
Method:	Least Squares	**Sparsity:**	0.09864
Date:	Mon, 27 Oct 2025	**No. Observations:**	38
Time:	08:58:23	**Df Residuals:**	36
		Df Model:	1

	coef	**std err**	**t**	**P> |t|**	**[0.025**	**0.975]**
Intercept	-0.0277	0.008	-3.444	0.001	-0.044	-0.011
sentiment_mean_top_influencer	0.0920	0.036	2.531	0.016	0.018	0.166

Table B.3. Quantile 0.5 — NVDA

Dep. Variable:	Return	**Pseudo R-squared:**	0.1484
Model:	QuantReg	**Bandwidth:**	0.04065
Method:	Least Squares	**Sparsity:**	0.08576
Date:	Mon, 27 Oct 2025	**No. Observations:**	38
Time:	08:58:23	**Df Residuals:**	36
		Df Model:	1

	coef	**std err**	**t**	**P> |t|**	**[0.025**	**0.975]**
Intercept	-0.0078	0.008	-1.016	0.317	-0.023	0.008
sentiment_mean_top_influencer	0.0721	0.028	2.580	0.014	0.015	0.129

Table B.4. Quantile 0.75 — NVDA

Dep. Variable:	Return	**Pseudo R-squared:**	0.1365
Model:	QuantReg	**Bandwidth:**	0.04412
Method:	Least Squares	**Sparsity:**	0.1017
Date:	Mon, 27 Oct 2025	**No. Observations:**	38
Time:	08:58:23	**Df Residuals:**	36
		Df Model:	1

	coef	**std err**	**t**	**P> |t|**	**[0.025**	**0.975]**
Intercept	0.0162	0.008	2.035	0.049	5.76e-05	0.032
sentiment_mean_top_influencer	0.0481	0.027	1.794	0.081	-0.006	0.103

Table B.5. Quantile 0.9 — NVDA

Dep. Variable:	Return	**Pseudo R-squared:**	0.1837
Model:	QuantReg	**Bandwidth:**	nan
Method:	Least Squares	**Sparsity:**	nan
Date:	Mon, 27 Oct 2025	**No. Observations:**	38
Time:	08:58:23	**Df Residuals:**	36
		Df Model:	1

	coef	**std err**	**t**	**P> |t|**	**[0.025**	**0.975]**
Intercept	0.0405	nan	nan	nan	nan	nan
sentiment_mean_top_influencer	0.0375	nan	nan	nan	nan	nan

Table B.6. OLS Regression Results for: NVDA

Dep. Variable:	Return	**R-squared:**	0.289
Model:	OLS	**Adj. R-squared:**	0.269
Method:	Least Squares	**F-statistic:**	14.64
Date:	Mon, 27 Oct 2025	**Prob (F-statistic):**	0.000500
Time:	08:58:23	**Log-Likelihood:**	81.517
No. Observations:	38	**AIC:**	-159.0
Df Residuals:	36	**BIC:**	-155.8
Df Model:	1		
Covariance Type:	nonrobust		

	coef	**std err**	**t**	**P> |t|**	**[0.025**	**0.975]**
Intercept	-0.0065	0.005	-1.240	0.223	-0.017	0.004
sentiment_mean_top_influencer	0.0725	0.019	3.826	0.000	0.034	0.111

Omnibus:	0.060	**Durbin-Watson:**	1.601
Prob(Omnibus):	0.971	**Jarque-Bera (JB):**	0.174
Skew:	-0.085	**Prob(JB):**	0.917
Kurtosis:	2.716	**Cond. No.**	4.08

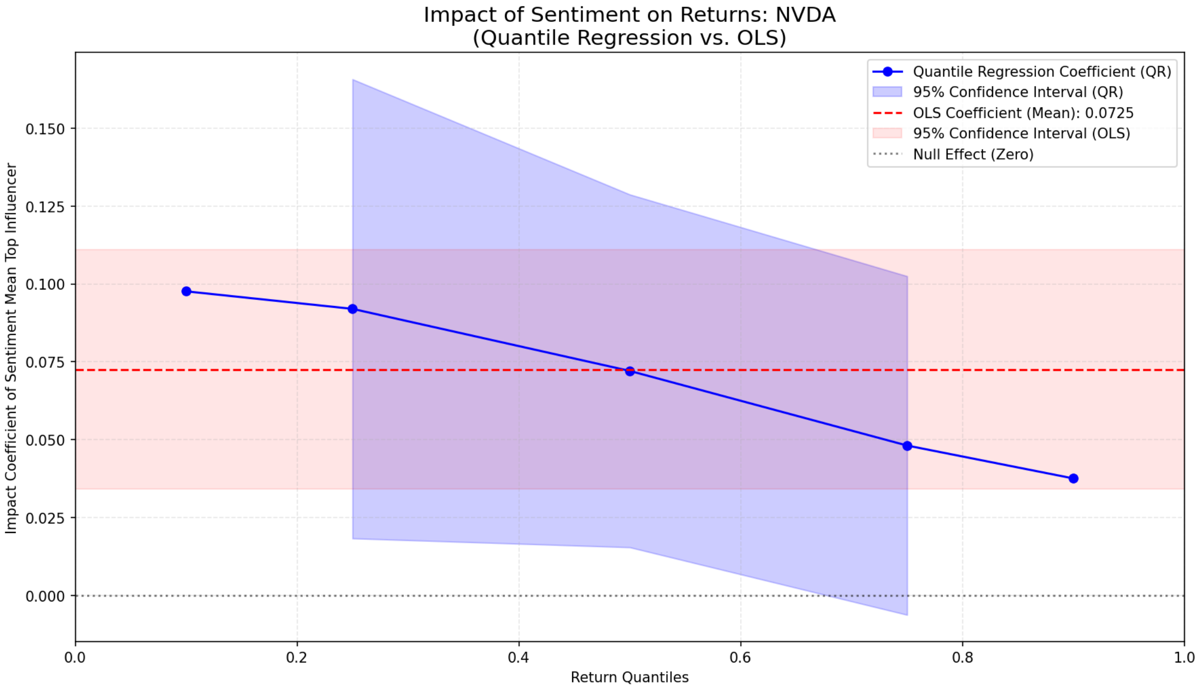

FIGURE B.1. Impact of Sentiment on Returns: NVDA (Quantile Regression vs. OLS)

Table B.7. Quantile 0.1 — BTC-USD

Dep. Variable:	Return	**Pseudo R-squared:**	0.09789
Model:	QuantReg	**Bandwidth:**	0.01726
Method:	Least Squares	**Sparsity:**	0.06497
Date:	Mon, 27 Oct 2025	**No. Observations:**	64
Time:	08:58:24	**Df Residuals:**	62
		Df Model:	1

	coef	**std err**	**t**	**P> |t|**	**[0.025**	**0.975]**
Intercept	-0.0210	0.004	-5.770	0.000	-0.028	-0.014
sentiment_mean_top_influencer	0.0314	0.018	1.772	0.081	-0.004	0.067

Table B.8. Quantile 0.25 — BTC-USD

Dep. Variable:	Return	**Pseudo R-squared:**	0.1663
Model:	QuantReg	**Bandwidth:**	0.01516
Method:	Least Squares	**Sparsity:**	0.04318
Date:	Mon, 27 Oct 2025	**No. Observations:**	64
Time:	08:58:24	**Df Residuals:**	62
		Df Model:	1

	coef	**std err**	**t**	**P> |t|**	**[0.025**	**0.975]**
Intercept	-0.0150	0.003	-4.775	0.000	-0.021	-0.009
sentiment_mean_top_influencer	0.0367	0.012	2.995	0.004	0.012	0.061

Table B.9. Quantile 0.5 — BTC-USD

Dep. Variable:	Return	**Pseudo R-squared:**	0.1005
Model:	QuantReg	**Bandwidth:**	0.01703
Method:	Least Squares	**Sparsity:**	0.03783
Date:	Mon, 27 Oct 2025	**No. Observations:**	64
Time:	08:58:24	**Df Residuals:**	62
		Df Model:	1

	coef	**std err**	**t**	**P> |t|**	**[0.025**	**0.975]**
Intercept	-0.0065	0.003	-2.168	0.034	-0.012	-0.001
sentiment_mean_top_influencer	0.0316	0.011	2.981	0.004	0.010	0.053

Table B.10. Quantile 0.75 — BTC-USD

Dep. Variable:	Return	**Pseudo R-squared:**	0.07718
Model:	QuantReg	**Bandwidth:**	0.01585
Method:	Least Squares	**Sparsity:**	0.05044
Date:	Mon, 27 Oct 2025	**No. Observations:**	64
Time:	08:58:24	**Df Residuals:**	62
		Df Model:	1

	coef	**std err**	**t**	**P> |t|**	**[0.025**	**0.975]**
Intercept	0.0044	0.004	1.249	0.216	-0.003	0.011
sentiment_mean_top_influencer	0.0263	0.013	2.045	0.045	0.001	0.052

Table B.11. Quantile 0.9 — BTC-USD

Dep. Variable:	Return	**Pseudo R-squared:**	0.04710
Model:	QuantReg	**Bandwidth:**	0.01503
Method:	Least Squares	**Sparsity:**	0.1799
Date:	Mon, 27 Oct 2025	**No. Observations:**	64
Time:	08:58:24	**Df Residuals:**	62
		Df Model:	1

	coef	std err	t	P> |t|	[0.025	0.975]
Intercept	0.0146	0.008	1.793	0.078	-0.002	0.031
sentiment_mean_top_influencer	0.0484	0.028	1.750	0.085	-0.007	0.104

Table B.12. OLS Regression Results for: BTC-USD

Dep. Variable:	Return	**R-squared:**	0.144
Model:	OLS	**Adj. R-squared:**	0.130
Method:	Least Squares	**F-statistic:**	10.42
Date:	Mon, 27 Oct 2025	**Prob (F-statistic):**	0.00199
Time:	08:58:24	**Log-Likelihood:**	162.63
No. Observations:	64	**AIC:**	-321.3
Df Residuals:	62	**BIC:**	-316.9
Df Model:	1		
Covariance Type:	nonrobust		

	coef	std err	t	P> |t|	[0.025	0.975]
Intercept	-0.0042	0.003	-1.372	0.175	-0.010	0.002
sentiment_mean_top_influencer	0.0350	0.011	3.229	0.002	0.013	0.057

Omnibus:	16.769	**Durbin-Watson:**	2.340
Prob(Omnibus):	0.000	**Jarque-Bera (JB):**	23.993
Skew:	0.985	**Prob(JB):**	6.16e-06
Kurtosis:	5.263	**Cond. No.**	4.62

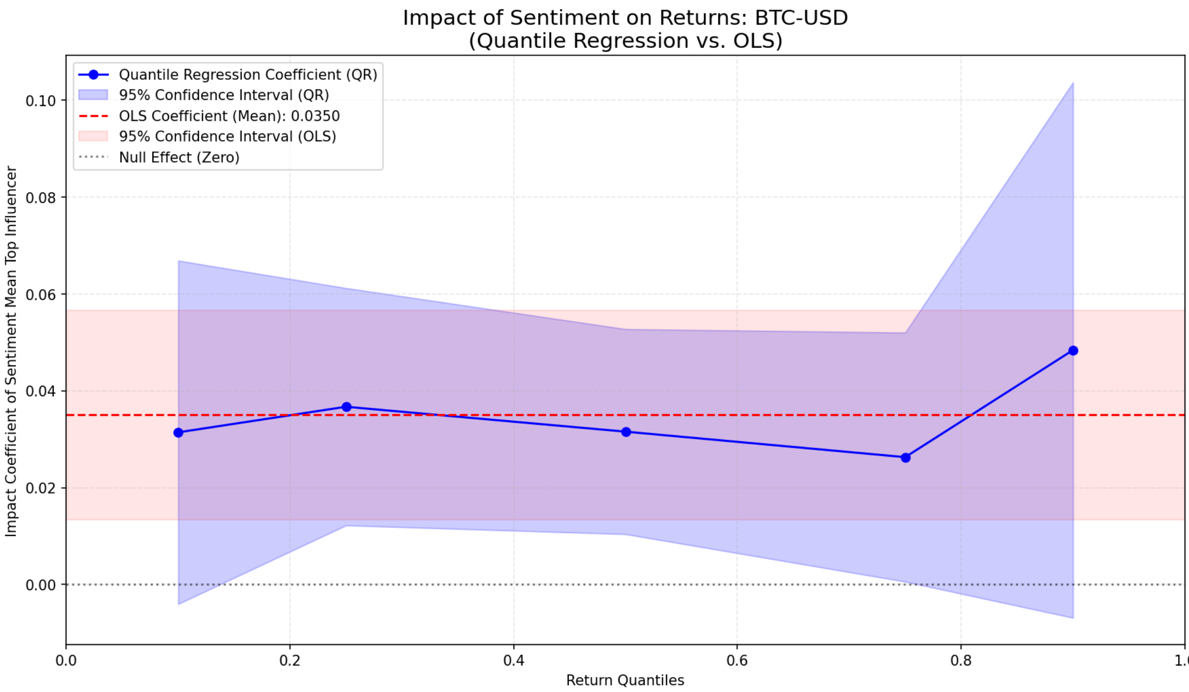

Figure B.2. Impact of Sentiment on Returns: BTC-USD (Quantile Regression vs. OLS)

Table B.13. Quantile 0.1 — META

Dep. Variable:	Return	**Pseudo R-squared:**	0.2926
Model:	QuantReg	**Bandwidth:**	0.01767
Method:	Least Squares	**Sparsity:**	0.1131
Date:	Mon, 27 Oct 2025	**No. Observations:**	82
Time:	08:58:24	**Df Residuals:**	80
		Df Model:	1

	coef	**std err**	**t**	**P> |t|**	**[0.025**	**0.975]**
Intercept	-0.0337	0.005	-7.406	0.000	-0.043	-0.025
sentiment_mean_top_influencer	0.0922	0.027	3.449	0.001	0.039	0.145

Table B.14. Quantile 0.25 — META

Dep. Variable:	Return	**Pseudo R-squared:**	0.2774
Model:	QuantReg	**Bandwidth:**	0.01694
Method:	Least Squares	**Sparsity:**	0.05171
Date:	Mon, 27 Oct 2025	**No. Observations:**	82
Time:	08:58:24	**Df Residuals:**	80
		Df Model:	1

	coef	**std err**	**t**	**P> |t|**	**[0.025**	**0.975]**
Intercept	-0.0182	0.003	-6.617	0.000	-0.024	-0.013
sentiment_mean_top_influencer	0.0786	0.014	5.495	0.000	0.050	0.107

Table B.15. Quantile 0.5 — META

Dep. Variable:	Return	**Pseudo R-squared:**	0.2487
Model:	QuantReg	**Bandwidth:**	0.01955
Method:	Least Squares	**Sparsity:**	0.04838
Date:	Mon, 27 Oct 2025	**No. Observations:**	82
Time:	08:58:24	**Df Residuals:**	80
		Df Model:	1

	coef	**std err**	**t**	**P> |t|**	**[0.025**	**0.975]**
Intercept	-0.0099	0.003	-3.420	0.001	-0.016	-0.004
sentiment_mean_top_influencer	0.0759	0.013	5.632	0.000	0.049	0.103

Table B.16. Quantile 0.75 — META

Dep. Variable:	Return	**Pseudo R-squared:**	0.1846
Model:	QuantReg	**Bandwidth:**	0.01705
Method:	Least Squares	**Sparsity:**	0.06806
Date:	Mon, 27 Oct 2025	**No. Observations:**	82
Time:	08:58:24	**Df Residuals:**	80
		Df Model:	1

	coef	**std err**	**t**	**P> |t|**	**[0.025**	**0.975]**
Intercept	0.0034	0.003	0.987	0.326	-0.003	0.010
sentiment_mean_top_influencer	0.0882	0.014	6.252	0.000	0.060	0.116

Table B.17. Quantile 0.9 — META

Dep. Variable:	Return	**Pseudo R-squared:**	0.07263
Model:	QuantReg	**Bandwidth:**	0.01902
Method:	Least Squares	**Sparsity:**	0.1551
Date:	Mon, 27 Oct 2025	**No. Observations:**	82
Time:	08:58:24	**Df Residuals:**	80
		Df Model:	1

	coef	**std err**	**t**	**P> |t|**	**[0.025**	**0.975]**
Intercept	0.0203	0.005	3.702	0.000	0.009	0.031
sentiment_mean_top_influencer	0.0729	0.022	3.345	0.001	0.030	0.116

Table B.18. OLS Regression Results for: META

Dep. Variable:	Return	**R-squared:**	0.265
Model:	OLS	**Adj. R-squared:**	0.255
Method:	Least Squares	**F-statistic:**	28.79
Date:	Mon, 27 Oct 2025	**Prob (F-statistic):**	7.67e-07
Time:	08:58:24	**Log-Likelihood:**	179.95
No. Observations:	82	**AIC:**	-355.9
Df Residuals:	80	**BIC:**	-351.1
Df Model:	1		
Covariance Type:	nonrobust		

	coef	**std err**	**t**	**P> |t|**	**[0.025**	**0.975]**
Intercept	-0.0067	0.003	-2.063	0.042	-0.013	-0.000
sentiment_mean_top_influencer	0.0816	0.015	5.365	0.000	0.051	0.112

Omnibus:	67.684	**Durbin-Watson:**	2.142
Prob(Omnibus):	0.000	**Jarque-Bera (JB):**	666.319
Skew:	2.288	**Prob(JB):**	2.04e-145
Kurtosis:	16.194	**Cond. No.**	5.08

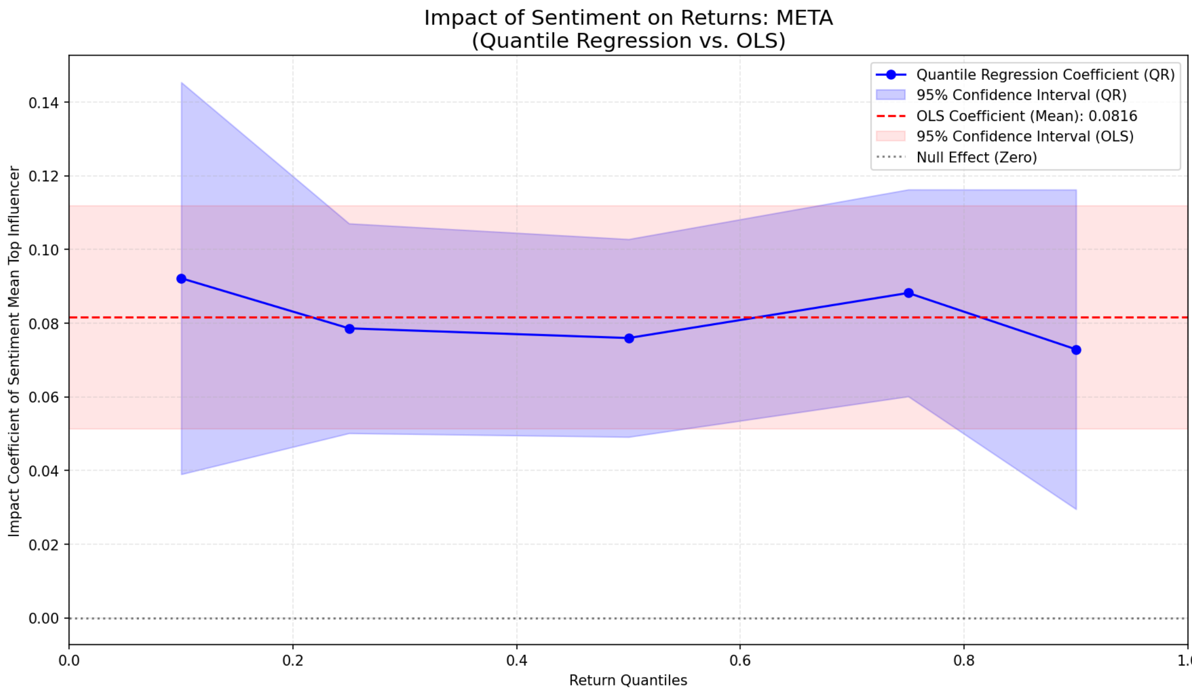

Figure B.3. Impact of Sentiment on Returns: META (Quantile Regression vs. OLS)

Table B.19. Quantile 0.1 — GOOGL

Dep. Variable:	Return	**Pseudo R-squared:**	0.1470
Model:	QuantReg	**Bandwidth:**	0.01627
Method:	Least Squares	**Sparsity:**	0.1269
Date:	Mon, 27 Oct 2025	**No. Observations:**	77
Time:	08:58:24	**Df Residuals:**	75
		Df Model:	1

	coef	**std err**	**t**	**P> |t|**	**[0.025**	**0.975]**
Intercept	-0.0290	0.006	-4.917	0.000	-0.041	-0.017
sentiment_mean_top_influencer	0.0578	0.028	2.099	0.039	0.003	0.113

Table B.20. Quantile 0.25 — GOOGL

Dep. Variable:	Return	**Pseudo R-squared:**	0.2139
Model:	QuantReg	**Bandwidth:**	0.01527
Method:	Least Squares	**Sparsity:**	0.04719
Date:	Mon, 27 Oct 2025	**No. Observations:**	77
Time:	08:58:24	**Df Residuals:**	75
		Df Model:	1

	coef	**std err**	**t**	**P> |t|**	**[0.025**	**0.975]**
Intercept	-0.0162	0.003	-5.756	0.000	-0.022	-0.011
sentiment_mean_top_influencer	0.0590	0.012	4.745	0.000	0.034	0.084

Table B.21. Quantile 0.5 — GOOGL

Dep. Variable:	Return	**Pseudo R-squared:**	0.2626
Model:	QuantReg	**Bandwidth:**	0.01625
Method:	Least Squares	**Sparsity:**	0.03913
Date:	Mon, 27 Oct 2025	**No. Observations:**	77
Time:	08:58:24	**Df Residuals:**	75
		Df Model:	1

	coef	**std err**	**t**	**P> |t|**	**[0.025**	**0.975]**
Intercept	-0.0076	0.003	-2.879	0.005	-0.013	-0.002
sentiment_mean_top_influencer	0.0672	0.011	5.904	0.000	0.045	0.090

Table B.22. Quantile 0.75 — GOOGL

Dep. Variable:	Return	**Pseudo R-squared:**	0.2036
Model:	QuantReg	**Bandwidth:**	0.01470
Method:	Least Squares	**Sparsity:**	0.04925
Date:	Mon, 27 Oct 2025	**No. Observations:**	77
Time:	08:58:24	**Df Residuals:**	75
		Df Model:	1

	coef	**std err**	**t**	**P> |t|**	**[0.025**	**0.975]**
Intercept	0.0011	0.003	0.388	0.699	-0.005	0.007
sentiment_mean_top_influencer	0.0666	0.012	5.437	0.000	0.042	0.091

Table B.23. Quantile 0.9 — GOOGL

Dep. Variable:	Return	**Pseudo R-squared:**	0.09572
Model:	QuantReg	**Bandwidth:**	0.01810
Method:	Least Squares	**Sparsity:**	0.09090
Date:	Mon, 27 Oct 2025	**No. Observations:**	77
Time:	08:58:24	**Df Residuals:**	75
		Df Model:	1

	coef	**std err**	**t**	**P> |t|**	**[0.025**	**0.975]**
Intercept	0.0155	0.004	4.336	0.000	0.008	0.023
sentiment_mean_top_influencer	0.0425	0.012	3.481	0.001	0.018	0.067

Table B.24. OLS Regression Results for: GOOGL

Dep. Variable:	Return	**R-squared:**	0.313
Model:	OLS	**Adj. R-squared:**	0.304
Method:	Least Squares	**F-statistic:**	34.25
Date:	Mon, 27 Oct 2025	**Prob (F-statistic):**	1.20e-07
Time:	08:58:24	**Log-Likelihood:**	192.62
No. Observations:	77	**AIC:**	-381.2
Df Residuals:	75	**BIC:**	-376.6
Df Model:	1		
Covariance Type:	nonrobust		

	coef	**std err**	**t**	**P> |t|**	**[0.025**	**0.975]**
Intercept	-0.0087	0.003	-3.182	0.002	-0.014	-0.003
sentiment_mean_top_influencer	0.0684	0.012	5.852	0.000	0.045	0.092

Omnibus:	21.530	**Durbin-Watson:**	2.200
Prob(Omnibus):	0.000	**Jarque-Bera (JB):**	87.492
Skew:	0.623	**Prob(JB):**	1.00e-19
Kurtosis:	8.071	**Cond. No.**	5.19

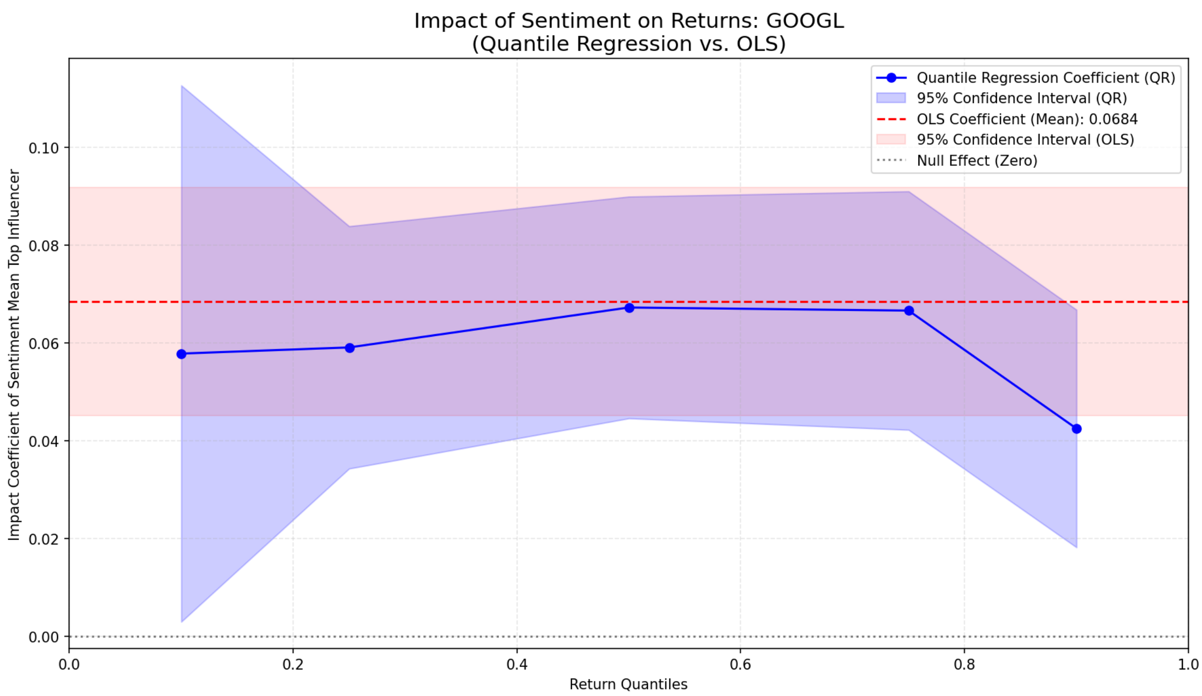

Figure B.4. Impact of Sentiment on Returns: GOOGL (Quantile Regression vs. OLS)

TABLE B.25. Quantile 0.1 — AMZN

Dep. Variable:	Return	**Pseudo R-squared:**	0.1833
Model:	QuantReg	**Bandwidth:**	0.02039
Method:	Least Squares	**Sparsity:**	0.08364
Date:	Mon, 27 Oct 2025	**No. Observations:**	80
Time:	08:58:25	**Df Residuals:**	78
		Df Model:	1

	coef	**std err**	**t**	**P> |t|**	**[0.025**	**0.975]**
Intercept	-0.0360	0.004	-9.477	0.000	-0.044	-0.028
sentiment_mean_top_influencer	0.0986	0.020	4.966	0.000	0.059	0.138

TABLE B.26. Quantile 0.25 — AMZN

Dep. Variable:	Return	**Pseudo R-squared:**	0.2264
Model:	QuantReg	**Bandwidth:**	0.01551
Method:	Least Squares	**Sparsity:**	0.04583
Date:	Mon, 27 Oct 2025	**No. Observations:**	80
Time:	08:58:25	**Df Residuals:**	78
		Df Model:	1

	coef	**std err**	**t**	**P> |t|**	**[0.025**	**0.975]**
Intercept	-0.0202	0.003	-7.086	0.000	-0.026	-0.015
sentiment_mean_top_influencer	0.0686	0.013	5.424	0.000	0.043	0.094

TABLE B.27. Quantile 0.5 — AMZN

Dep. Variable:	Return	**Pseudo R-squared:**	0.1835
Model:	QuantReg	**Bandwidth:**	0.01692
Method:	Least Squares	**Sparsity:**	0.04324
Date:	Mon, 27 Oct 2025	**No. Observations:**	80
Time:	08:58:25	**Df Residuals:**	78
		Df Model:	1

	coef	**std err**	**t**	**P> |t|**	**[0.025**	**0.975]**
Intercept	-0.0142	0.003	-4.655	0.000	-0.020	-0.008
sentiment_mean_top_influencer	0.0674	0.013	5.049	0.000	0.041	0.094

TABLE B.28. Quantile 0.75 — AMZN

Dep. Variable:	Return	**Pseudo R-squared:**	0.08950
Model:	QuantReg	**Bandwidth:**	0.01646
Method:	Least Squares	**Sparsity:**	0.06619
Date:	Mon, 27 Oct 2025	**No. Observations:**	80
Time:	08:58:25	**Df Residuals:**	78
		Df Model:	1

	coef	**std err**	**t**	**P> |t|**	**[0.025**	**0.975]**
Intercept	0.0015	0.004	0.360	0.720	-0.007	0.010
sentiment_mean_top_influencer	0.0561	0.019	3.031	0.003	0.019	0.093

Table B.29. Quantile 0.9 — AMZN

Dep. Variable:	Return	**Pseudo R-squared:**	0.03250
Model:	QuantReg	**Bandwidth:**	0.02024
Method:	Least Squares	**Sparsity:**	0.1240
Date:	Mon, 27 Oct 2025	**No. Observations:**	80
Time:	08:58:25	**Df Residuals:**	78
		Df Model:	1

	coef	**std err**	**t**	**P> |t|**	**[0.025**	**0.975]**
Intercept	0.0168	0.005	3.220	0.002	0.006	0.027
sentiment_mean_top_influencer	0.0440	0.023	1.873	0.065	-0.003	0.091

Table B.30. OLS Regression Results for: AMZN

Dep. Variable:	Return	**R-squared:**	0.169
Model:	OLS	**Adj. R-squared:**	0.158
Method:	Least Squares	**F-statistic:**	15.83
Date:	Mon, 27 Oct 2025	**Prob (F-statistic):**	0.000154
Time:	08:58:25	**Log-Likelihood:**	181.54
No. Observations:	80	**AIC:**	-359.1
Df Residuals:	78	**BIC:**	-354.3
Df Model:	1		
Covariance Type:	nonrobust		

	coef	**std err**	**t**	**P> |t|**	**[0.025**	**0.975]**
Intercept	-0.0088	0.004	-2.444	0.017	-0.016	-0.002
sentiment_mean_top_influencer	0.0622	0.016	3.979	0.000	0.031	0.093

Omnibus:	45.346	**Durbin-Watson:**	2.187
Prob(Omnibus):	0.000	**Jarque-Bera (JB):**	288.075
Skew:	1.481	**Prob(JB):**	2.79e-63
Kurtosis:	11.812	**Cond. No.**	5.63

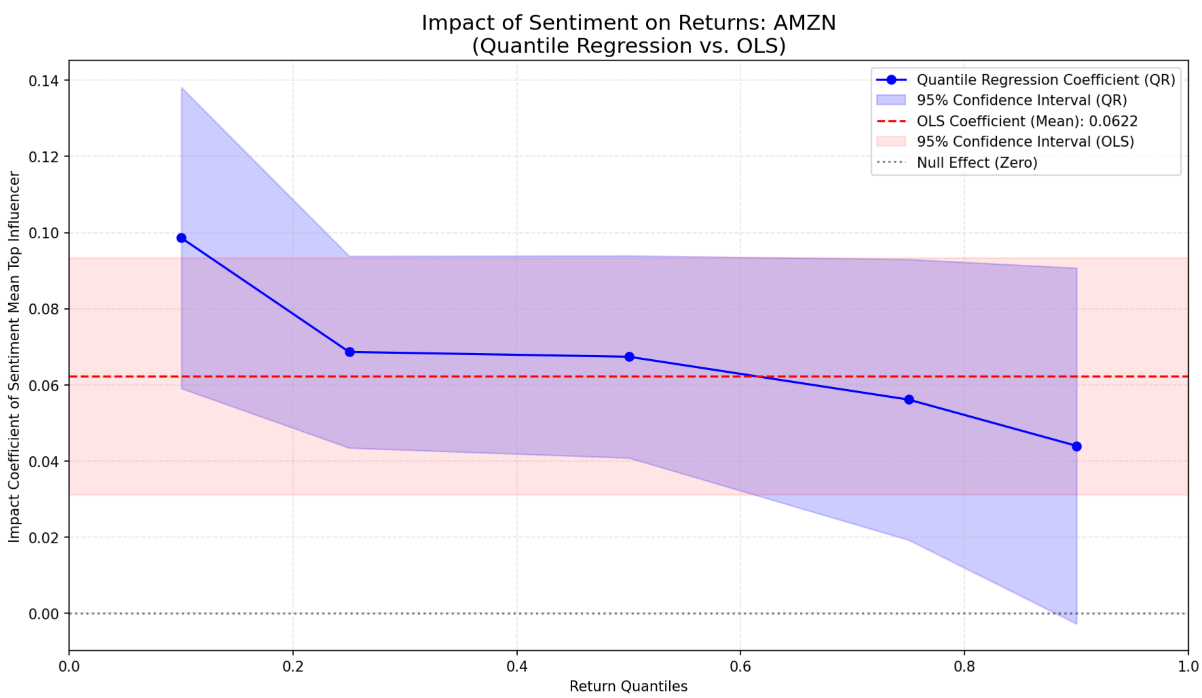

Figure B.5. Impact of Sentiment on Returns: AMZN (Quantile Regression vs. OLS)

Table B.31. Quantile 0.1 — NIO

Dep. Variable:	Return	**Pseudo R-squared:**	0.2112			
Model:	QuantReg	**Bandwidth:**	0.03496			
Method:	Least Squares	**Sparsity:**	0.1499			
Date:	Mon, 27 Oct 2025	**No. Observations:**	67			
Time:	08:58:25	**Df Residuals:**	65			
		Df Model:	1			
	coef	**std err**	**t**	**P> |t|**	**[0.025**	**0.975]**
Intercept	-0.0604	0.008	-7.204	0.000	-0.077	-0.044
sentiment_mean_top_influencer	0.1055	0.034	3.104	0.003	0.038	0.173

Table B.32. Quantile 0.25 — NIO

Dep. Variable:	Return	**Pseudo R-squared:**	0.1402			
Model:	QuantReg	**Bandwidth:**	0.03405			
Method:	Least Squares	**Sparsity:**	0.1076			
Date:	Mon, 27 Oct 2025	**No. Observations:**	67			
Time:	08:58:25	**Df Residuals:**	65			
		Df Model:	1			
	coef	**std err**	**t**	**P> |t|**	**[0.025**	**0.975]**
Intercept	-0.0561	0.009	-6.040	0.000	-0.075	-0.038
sentiment_mean_top_influencer	0.1407	0.036	3.931	0.000	0.069	0.212

Table B.33. Quantile 0.5 — NIO

Dep. Variable:	Return	**Pseudo R-squared:**	0.2275			
Model:	QuantReg	**Bandwidth:**	0.03797			
Method:	Least Squares	**Sparsity:**	0.08133			
Date:	Mon, 27 Oct 2025	**No. Observations:**	67			
Time:	08:58:25	**Df Residuals:**	65			
		Df Model:	1			
	coef	**std err**	**t**	**P> |t|**	**[0.025**	**0.975]**
Intercept	-0.0334	0.008	-4.146	0.000	-0.049	-0.017
sentiment_mean_top_influencer	0.1484	0.032	4.648	0.000	0.085	0.212

Table B.34. Quantile 0.75 — NIO

Dep. Variable:	Return	**Pseudo R-squared:**	0.2826			
Model:	QuantReg	**Bandwidth:**	0.03351			
Method:	Least Squares	**Sparsity:**	0.08754			
Date:	Mon, 27 Oct 2025	**No. Observations:**	67			
Time:	08:58:25	**Df Residuals:**	65			
		Df Model:	1			
	coef	**std err**	**t**	**P> |t|**	**[0.025**	**0.975]**
Intercept	-0.0182	0.008	-2.266	0.027	-0.034	-0.002
sentiment_mean_top_influencer	0.1513	0.032	4.685	0.000	0.087	0.216

Table B.35. Quantile 0.9 — NIO

Dep. Variable:	Return	**Pseudo R-squared:**	0.1974
Model:	QuantReg	**Bandwidth:**	0.03153
Method:	Least Squares	**Sparsity:**	0.1573
Date:	Mon, 27 Oct 2025	**No. Observations:**	67
Time:	08:58:25	**Df Residuals:**	65
		Df Model:	1

	coef	std err	t	P> |t|	[0.025	0.975]
Intercept	-0.0098	0.009	-1.036	0.304	-0.029	0.009
sentiment_mean_top_influencer	0.1981	0.036	5.488	0.000	0.126	0.270

Table B.36. OLS Regression Results for: NIO

Dep. Variable:	Return	**R-squared:**	0.369
Model:	OLS	**Adj. R-squared:**	0.359
Method:	Least Squares	**F-statistic:**	37.93
Date:	Mon, 27 Oct 2025	**Prob (F-statistic):**	5.15e-08
Time:	08:58:25	**Log-Likelihood:**	133.06
No. Observations:	67	**AIC:**	-262.1
Df Residuals:	65	**BIC:**	-257.7
Df Model:	1		
Covariance Type:	nonrobust		

	coef	std err	t	P> |t|	[0.025	0.975]
Intercept	-0.0356	0.007	-5.328	0.000	-0.049	-0.022
sentiment_mean_top_influencer	0.1631	0.026	6.159	0.000	0.110	0.216

Omnibus:	20.109	**Durbin-Watson:**	2.091
Prob(Omnibus):	0.000	**Jarque-Bera (JB):**	36.660
Skew:	1.014	**Prob(JB):**	1.10e-08
Kurtosis:	6.003	**Cond. No.**	6.69

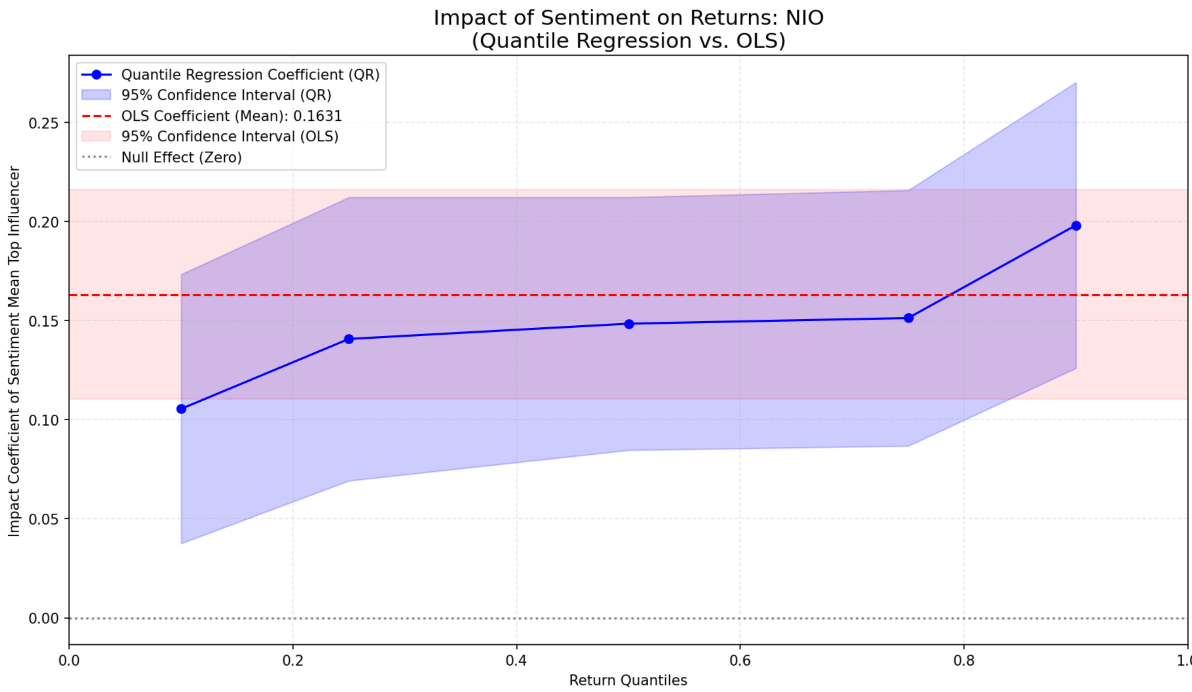

Figure B.6. Impact of Sentiment on Returns: NIO (Quantile Regression vs. OLS)

Table B.37. Quantile 0.1 — MSFT

Dep. Variable:	Return	**Pseudo R-squared:**	0.2415
Model:	QuantReg	**Bandwidth:**	0.01562
Method:	Least Squares	**Sparsity:**	0.05524
Date:	Mon, 27 Oct 2025	**No. Observations:**	82
Time:	08:58:26	**Df Residuals:**	80
		Df Model:	1

	coef	**std err**	**t**	**P> |t|**	**[0.025**	**0.975]**
Intercept	-0.0240	0.003	-7.810	0.000	-0.030	-0.018
sentiment_mean_top_influencer	0.0617	0.017	3.635	0.000	0.028	0.095

Table B.38. Quantile 0.25 — MSFT

Dep. Variable:	Return	**Pseudo R-squared:**	0.1731
Model:	QuantReg	**Bandwidth:**	0.01487
Method:	Least Squares	**Sparsity:**	0.04113
Date:	Mon, 27 Oct 2025	**No. Observations:**	82
Time:	08:58:26	**Df Residuals:**	80
		Df Model:	1

	coef	**std err**	**t**	**P> |t|**	**[0.025**	**0.975]**
Intercept	-0.0170	0.003	-5.612	0.000	-0.023	-0.011
sentiment_mean_top_influencer	0.0565	0.014	3.964	0.000	0.028	0.085

Table B.39. Quantile 0.5 — MSFT

Dep. Variable:	Return	**Pseudo R-squared:**	0.09640
Model:	QuantReg	**Bandwidth:**	0.01488
Method:	Least Squares	**Sparsity:**	0.03602
Date:	Mon, 27 Oct 2025	**No. Observations:**	82
Time:	08:58:26	**Df Residuals:**	80
		Df Model:	1

	coef	**std err**	**t**	**P> |t|**	**[0.025**	**0.975]**
Intercept	-0.0043	0.003	-1.535	0.129	-0.010	0.001
sentiment_mean_top_influencer	0.0332	0.012	2.666	0.009	0.008	0.058

Table B.40. Quantile 0.75 — MSFT

Dep. Variable:	Return	**Pseudo R-squared:**	0.07394
Model:	QuantReg	**Bandwidth:**	0.01399
Method:	Least Squares	**Sparsity:**	0.04488
Date:	Mon, 27 Oct 2025	**No. Observations:**	82
Time:	08:58:26	**Df Residuals:**	80
		Df Model:	1

	coef	**std err**	**t**	**P> |t|**	**[0.025**	**0.975]**
Intercept	0.0024	0.003	0.785	0.435	-0.004	0.009
sentiment_mean_top_influencer	0.0406	0.013	3.022	0.003	0.014	0.067

TABLE B.41. Quantile 0.9 — MSFT

Dep. Variable:	Return	**Pseudo R-squared:**	0.1156
Model:	QuantReg	**Bandwidth:**	0.01571
Method:	Least Squares	**Sparsity:**	0.08594
Date:	Mon, 27 Oct 2025	**No. Observations:**	82
Time:	08:58:26	**Df Residuals:**	80
		Df Model:	1

	coef	**std err**	**t**	**P> |t|**	**[0.025**	**0.975]**
Intercept	0.0107	0.004	2.422	0.018	0.002	0.019
sentiment_mean_top_influencer	0.0562	0.020	2.881	0.005	0.017	0.095

TABLE B.42. OLS Regression Results for: MSFT

Dep. Variable:	Return	**R-squared:**	0.213
Model:	OLS	**Adj. R-squared:**	0.203
Method:	Least Squares	**F-statistic:**	21.64
Date:	Mon, 27 Oct 2025	**Prob (F-statistic):**	1.28e-05
Time:	08:58:26	**Log-Likelihood:**	214.37
No. Observations:	82	**AIC:**	-424.7
Df Residuals:	80	**BIC:**	-419.9
Df Model:	1		
Covariance Type:	nonrobust		

	coef	**std err**	**t**	**P> |t|**	**[0.025**	**0.975]**
Intercept	-0.0069	0.003	-2.488	0.015	-0.012	-0.001
sentiment_mean_top_influencer	0.0578	0.012	4.652	0.000	0.033	0.082

Omnibus:	55.171	**Durbin-Watson:**	2.035
Prob(Omnibus):	0.000	**Jarque-Bera (JB):**	270.440
Skew:	2.045	**Prob(JB):**	1.88e-59
Kurtosis:	10.901	**Cond. No.**	6.43

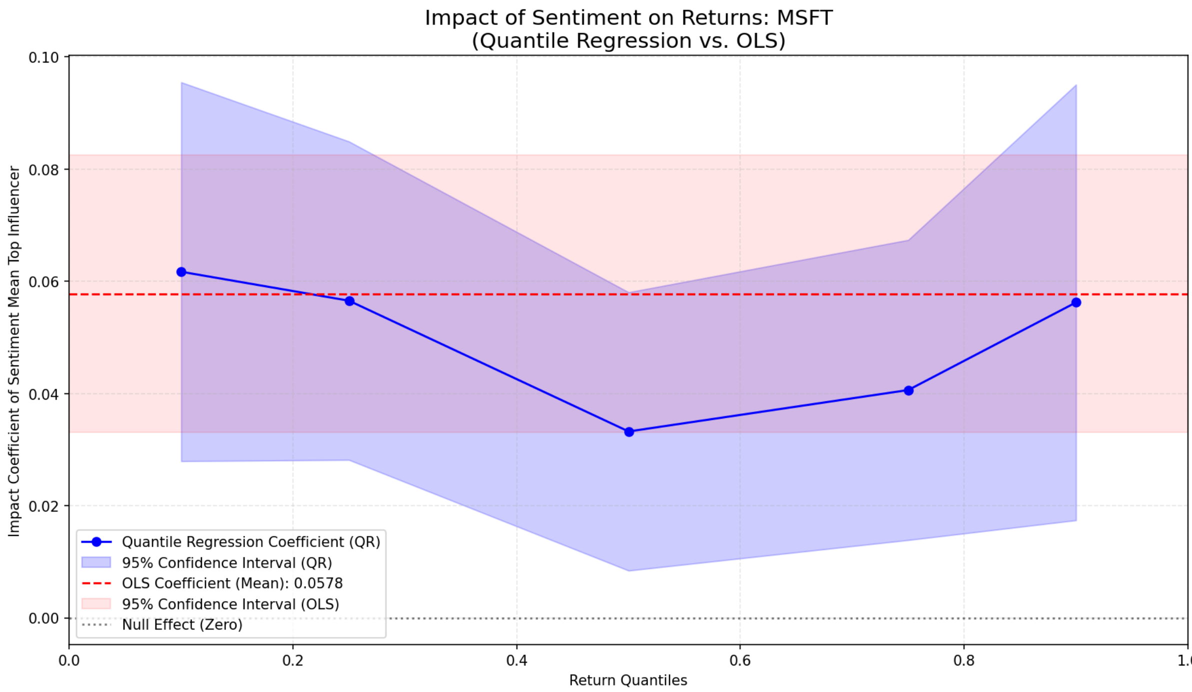

FIGURE B.7. Impact of Sentiment on Returns: MSFT (Quantile Regression vs. OLS)

Table B.43. Quantile 0.1 — INTC

Dep. Variable:	Return	**Pseudo R-squared:**	0.2085
Model:	QuantReg	**Bandwidth:**	0.02757
Method:	Least Squares	**Sparsity:**	0.1180
Date:	Mon, 27 Oct 2025	**No. Observations:**	65
Time:	08:58:26	**Df Residuals:**	63
		Df Model:	1

	coef	**std err**	**t**	**P> |t|**	**[0.025**	**0.975]**
Intercept	-0.0597	0.007	-8.757	0.000	-0.073	-0.046
sentiment_mean_top_influencer	0.1899	0.036	5.323	0.000	0.119	0.261

Table B.44. Quantile 0.25 — INTC

Dep. Variable:	Return	**Pseudo R-squared:**	0.2386
Model:	QuantReg	**Bandwidth:**	0.02398
Method:	Least Squares	**Sparsity:**	0.06897
Date:	Mon, 27 Oct 2025	**No. Observations:**	65
Time:	08:58:26	**Df Residuals:**	63
		Df Model:	1

	coef	**std err**	**t**	**P> |t|**	**[0.025**	**0.975]**
Intercept	-0.0358	0.005	-6.740	0.000	-0.046	-0.025
sentiment_mean_top_influencer	0.1413	0.023	6.279	0.000	0.096	0.186

Table B.45. Quantile 0.5 — INTC

Dep. Variable:	Return	**Pseudo R-squared:**	0.2344
Model:	QuantReg	**Bandwidth:**	0.02476
Method:	Least Squares	**Sparsity:**	0.05686
Date:	Mon, 27 Oct 2025	**No. Observations:**	65
Time:	08:58:26	**Df Residuals:**	63
		Df Model:	1

	coef	**std err**	**t**	**P> |t|**	**[0.025**	**0.975]**
Intercept	-0.0205	0.006	-3.708	0.000	-0.032	-0.009
sentiment_mean_top_influencer	0.1259	0.025	4.939	0.000	0.075	0.177

Table B.46. Quantile 0.75 — INTC

Dep. Variable:	Return	**Pseudo R-squared:**	0.2371
Model:	QuantReg	**Bandwidth:**	0.02355
Method:	Least Squares	**Sparsity:**	0.07390
Date:	Mon, 27 Oct 2025	**No. Observations:**	65
Time:	08:58:26	**Df Residuals:**	63
		Df Model:	1

	coef	**std err**	**t**	**P> |t|**	**[0.025**	**0.975]**
Intercept	-0.0100	0.006	-1.544	0.128	-0.023	0.003
sentiment_mean_top_influencer	0.1477	0.031	4.811	0.000	0.086	0.209

TABLE B.47. Quantile 0.9 — INTC

Dep. Variable:	Return	**Pseudo R-squared:**	0.2174
Model:	QuantReg	**Bandwidth:**	0.02760
Method:	Least Squares	**Sparsity:**	0.1158
Date:	Mon, 27 Oct 2025	**No. Observations:**	65
Time:	08:58:26	**Df Residuals:**	63
		Df Model:	1

	coef	std err	t	P> |t|	[0.025	0.975]
Intercept	-0.0032	0.008	-0.396	0.693	-0.019	0.013
sentiment_mean_top_influencer	0.1879	0.039	4.852	0.000	0.110	0.265

TABLE B.48. OLS Regression Results for: INTC

Dep. Variable:	Return	**R-squared:**	0.310
Model:	OLS	**Adj. R-squared:**	0.299
Method:	Least Squares	**F-statistic:**	28.36
Date:	Mon, 27 Oct 2025	**Prob (F-statistic):**	1.43e-06
Time:	08:58:26	**Log-Likelihood:**	125.59
No. Observations:	65	**AIC:**	-247.2
Df Residuals:	63	**BIC:**	-242.8
Df Model:	1		
Covariance Type:	nonrobust		

	coef	std err	t	P> |t|	[0.025	0.975]
Intercept	-0.0251	0.007	-3.618	0.001	-0.039	-0.011
sentiment_mean_top_influencer	0.1699	0.032	5.325	0.000	0.106	0.234

Omnibus:	50.617	**Durbin-Watson:**	2.092
Prob(Omnibus):	0.000	**Jarque-Bera (JB):**	253.354
Skew:	2.171	**Prob(JB):**	9.66e-56
Kurtosis:	11.642	**Cond. No.**	7.43

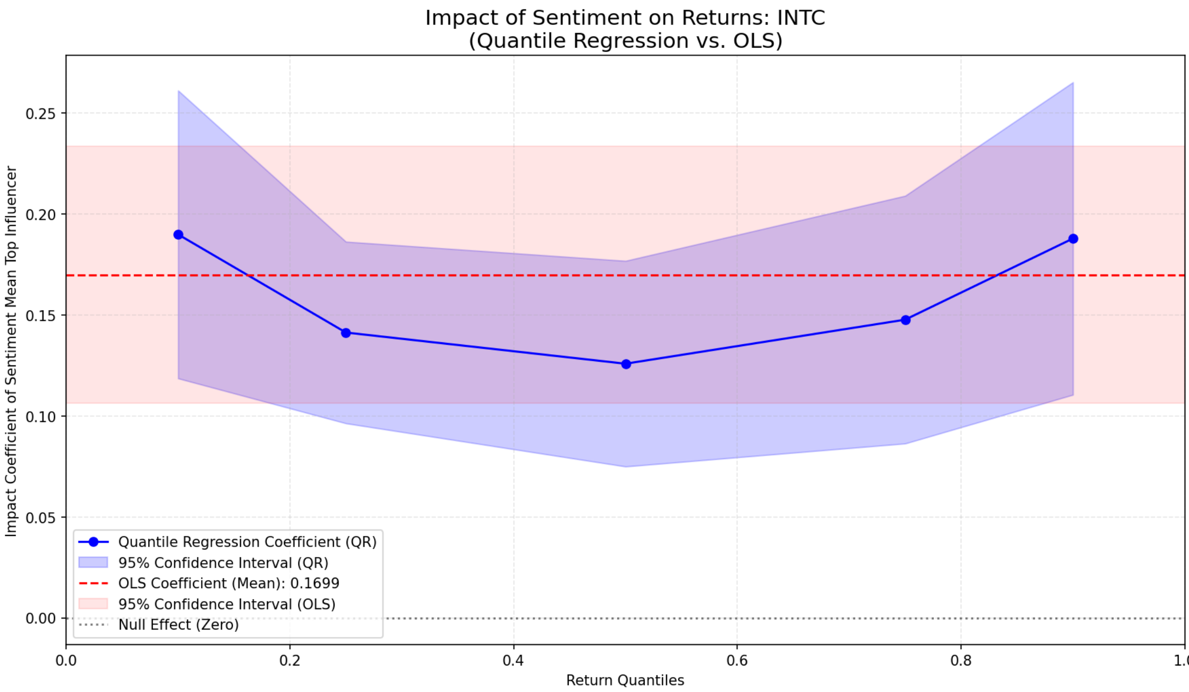

FIGURE B.8. Impact of Sentiment on Returns: INTC (Quantile Regression vs. OLS)

TABLE B.49. Quantile 0.1 — AMD

Dep. Variable:	Return	**Pseudo R-squared:**	0.2553
Model:	QuantReg	**Bandwidth:**	0.03478
Method:	Least Squares	**Sparsity:**	0.1246
Date:	Mon, 27 Oct 2025	**No. Observations:**	66
Time:	08:58:27	**Df Residuals:**	64
		Df Model:	1

	coef	**std err**	**t**	**P> |t|**	**[0.025**	**0.975]**
Intercept	-0.0502	0.008	-6.229	0.000	-0.066	-0.034
sentiment_mean_top_influencer	0.0984	0.032	3.082	0.003	0.035	0.162

TABLE B.50. Quantile 0.25 — AMD

Dep. Variable:	Return	**Pseudo R-squared:**	0.1404
Model:	QuantReg	**Bandwidth:**	0.03297
Method:	Least Squares	**Sparsity:**	0.08558
Date:	Mon, 27 Oct 2025	**No. Observations:**	66
Time:	08:58:27	**Df Residuals:**	64
		Df Model:	1

	coef	**std err**	**t**	**P> |t|**	**[0.025**	**0.975]**
Intercept	-0.0312	0.007	-4.589	0.000	-0.045	-0.018
sentiment_mean_top_influencer	0.0786	0.027	2.958	0.004	0.026	0.132

TABLE B.51. Quantile 0.5 — AMD

Dep. Variable:	Return	**Pseudo R-squared:**	0.1180
Model:	QuantReg	**Bandwidth:**	0.03469
Method:	Least Squares	**Sparsity:**	0.08132
Date:	Mon, 27 Oct 2025	**No. Observations:**	66
Time:	08:58:27	**Df Residuals:**	64
		Df Model:	1

	coef	**std err**	**t**	**P> |t|**	**[0.025**	**0.975]**
Intercept	-0.0139	0.007	-2.017	0.048	-0.028	-0.000
sentiment_mean_top_influencer	0.0708	0.025	2.803	0.007	0.020	0.121

TABLE B.52. Quantile 0.75 — AMD

Dep. Variable:	Return	**Pseudo R-squared:**	0.1289
Model:	QuantReg	**Bandwidth:**	0.03135
Method:	Least Squares	**Sparsity:**	0.09443
Date:	Mon, 27 Oct 2025	**No. Observations:**	66
Time:	08:58:27	**Df Residuals:**	64
		Df Model:	1

	coef	**std err**	**t**	**P> |t|**	**[0.025**	**0.975]**
Intercept	0.0013	0.007	0.193	0.848	-0.012	0.015
sentiment_mean_top_influencer	0.0941	0.026	3.619	0.001	0.042	0.146

Table B.53. Quantile 0.9 — AMD

Dep. Variable:	Return	**Pseudo R-squared:**	0.1115
Model:	QuantReg	**Bandwidth:**	0.03611
Method:	Least Squares	**Sparsity:**	0.1522
Date:	Mon, 27 Oct 2025	**No. Observations:**	66
Time:	08:58:27	**Df Residuals:**	64
		Df Model:	1

	coef	**std err**	**t**	**P> |t|**	**[0.025**	**0.975]**
Intercept	0.0226	0.009	2.641	0.010	0.006	0.040
sentiment_mean_top_influencer	0.0779	0.032	2.432	0.018	0.014	0.142

Table B.54. OLS Regression Results for: AMD

Dep. Variable:	Return	**R-squared:**	0.159
Model:	OLS	**Adj. R-squared:**	0.146
Method:	Least Squares	**F-statistic:**	12.08
Date:	Mon, 27 Oct 2025	**Prob (F-statistic):**	0.000922
Time:	08:58:27	**Log-Likelihood:**	120.99
No. Observations:	66	**AIC:**	-238.0
Df Residuals:	64	**BIC:**	-233.6
Df Model:	1		
Covariance Type:	nonrobust		

	coef	**std err**	**t**	**P> |t|**	**[0.025**	**0.975]**
Intercept	-0.0123	0.007	-1.847	0.069	-0.026	0.001
sentiment_mean_top_influencer	0.0849	0.024	3.475	0.001	0.036	0.134

Omnibus:	80.502	**Durbin-Watson:**	2.502
Prob(Omnibus):	0.000	**Jarque-Bera (JB):**	1133.310
Skew:	3.401	**Prob(JB):**	8.03e-247
Kurtosis:	22.127	**Cond. No.**	5.24

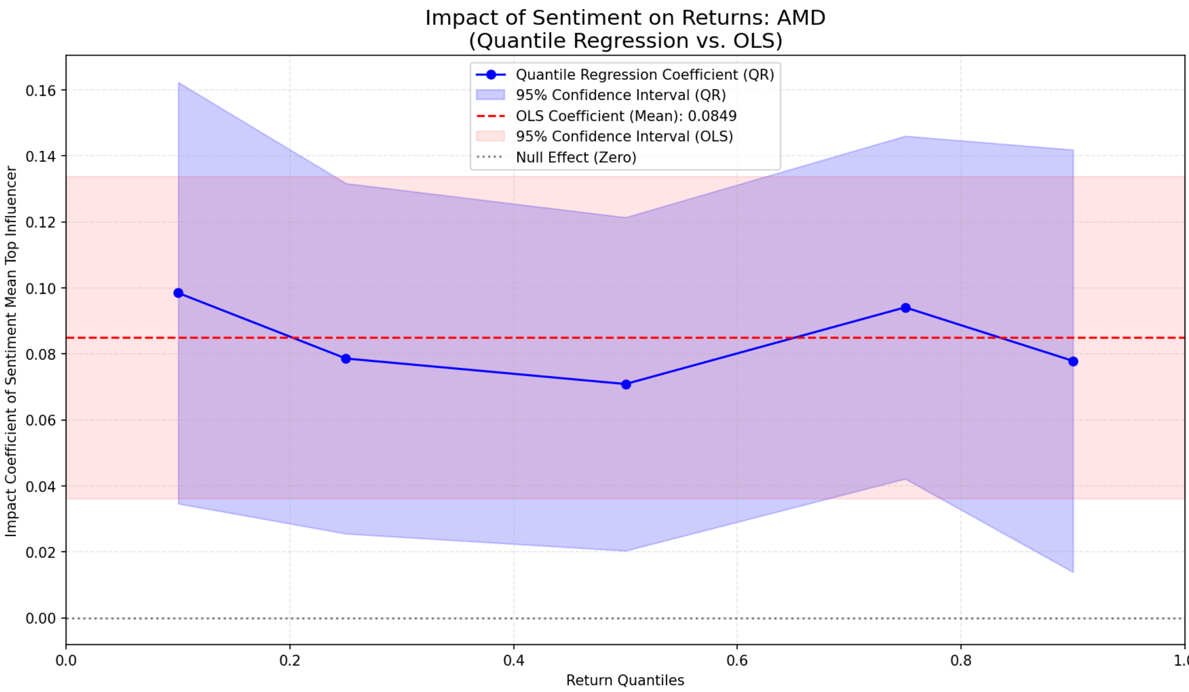

Figure B.9. Impact of Sentiment on Returns: AMD (Quantile Regression vs. OLS)

Table B.55. Quantile 0.1 — XRP-USD

Dep. Variable:	Return	**Pseudo R-squared:**	0.1906
Model:	QuantReg	**Bandwidth:**	0.02591
Method:	Least Squares	**Sparsity:**	0.1741
Date:	Mon, 27 Oct 2025	**No. Observations:**	78
Time:	08:58:27	**Df Residuals:**	76
		Df Model:	1

	coef	std err	t	P> |t|	[0.025	0.975]
Intercept	-0.0785	0.024	-3.299	0.001	-0.126	-0.031
sentiment_mean_top_influencer	0.1038	0.060	1.729	0.088	-0.016	0.223

Table B.56. Quantile 0.25 — XRP-USD

Dep. Variable:	Return	**Pseudo R-squared:**	0.08568
Model:	QuantReg	**Bandwidth:**	0.02159
Method:	Least Squares	**Sparsity:**	0.08811
Date:	Mon, 27 Oct 2025	**No. Observations:**	78
Time:	08:58:27	**Df Residuals:**	76
		Df Model:	1

	coef	std err	t	P> |t|	[0.025	0.975]
Intercept	-0.0448	0.013	-3.565	0.001	-0.070	-0.020
sentiment_mean_top_influencer	0.0737	0.031	2.397	0.019	0.012	0.135

Table B.57. Quantile 0.5 — XRP-USD

Dep. Variable:	Return	**Pseudo R-squared:**	0.1269
Model:	QuantReg	**Bandwidth:**	0.02511
Method:	Least Squares	**Sparsity:**	0.06428
Date:	Mon, 27 Oct 2025	**No. Observations:**	78
Time:	08:58:27	**Df Residuals:**	76
		Df Model:	1

	coef	std err	t	P> |t|	[0.025	0.975]
Intercept	-0.0372	0.010	-3.876	0.000	-0.056	-0.018
sentiment_mean_top_influencer	0.0943	0.023	4.082	0.000	0.048	0.140

Table B.58. Quantile 0.75 — XRP-USD

Dep. Variable:	Return	**Pseudo R-squared:**	0.1614
Model:	QuantReg	**Bandwidth:**	0.02513
Method:	Least Squares	**Sparsity:**	0.06934
Date:	Mon, 27 Oct 2025	**No. Observations:**	78
Time:	08:58:27	**Df Residuals:**	76
		Df Model:	1

	coef	std err	t	P> |t|	[0.025	0.975]
Intercept	-0.0315	0.009	-3.369	0.001	-0.050	-0.013
sentiment_mean_top_influencer	0.1155	0.023	5.041	0.000	0.070	0.161

TABLE B.59. Quantile 0.9 — XRP-USD

Dep. Variable:	Return	**Pseudo R-squared:**	0.1726
Model:	QuantReg	**Bandwidth:**	0.02799
Method:	Least Squares	**Sparsity:**	0.1249
Date:	Mon, 27 Oct 2025	**No. Observations:**	78
Time:	08:58:27	**Df Residuals:**	76
		Df Model:	1

	coef	std err	t	P> |t|	[0.025	0.975]
Intercept	-0.0346	0.013	-2.749	0.007	-0.060	-0.010
sentiment_mean_top_influencer	0.1653	0.031	5.331	0.000	0.104	0.227

TABLE B.60. OLS Regression Results for: XRP-USD

Dep. Variable:	Return	**R-squared:**	0.270
Model:	OLS	**Adj. R-squared:**	0.260
Method:	Least Squares	**F-statistic:**	28.10
Date:	Mon, 27 Oct 2025	**Prob (F-statistic):**	1.09e-06
Time:	08:58:27	**Log-Likelihood:**	166.13
No. Observations:	78	**AIC:**	-328.3
Df Residuals:	76	**BIC:**	-323.5
Df Model:	1		
Covariance Type:	nonrobust		

	coef	std err	t	P> |t|	[0.025	0.975]
Intercept	-0.0429	0.009	-4.928	0.000	-0.060	-0.026
sentiment_mean_top_influencer	0.1111	0.021	5.301	0.000	0.069	0.153

Omnibus:	11.658	**Durbin-Watson:**	1.822
Prob(Omnibus):	0.003	**Jarque-Bera (JB):**	16.636
Skew:	0.608	**Prob(JB):**	0.000244
Kurtosis:	4.907	**Cond. No.**	7.31

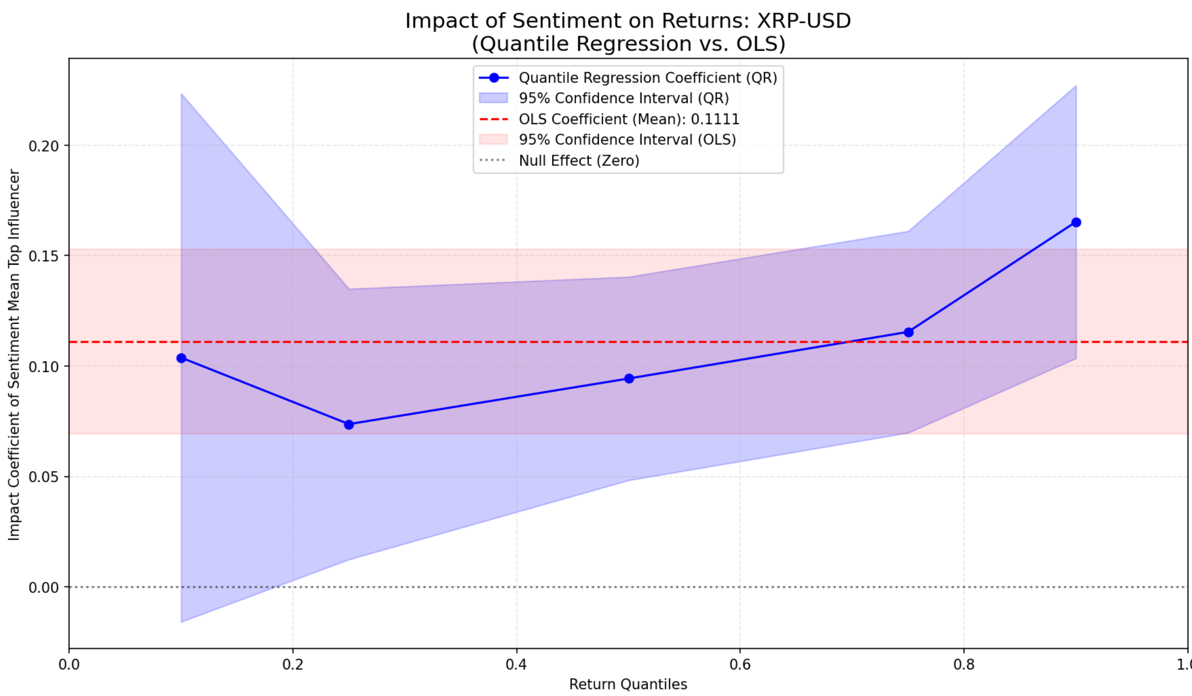

FIGURE B.10. Impact of Sentiment on Returns: XRP-USD (Quantile Regression vs. OLS)

Table B.61. Quantile 0.1 — LCID

Dep. Variable:	Return	**Pseudo R-squared:**	0.1438
Model:	QuantReg	**Bandwidth:**	0.04062
Method:	Least Squares	**Sparsity:**	0.1797
Date:	Mon, 27 Oct 2025	**No. Observations:**	68
Time:	08:58:28	**Df Residuals:**	66
		Df Model:	1

	coef	**std err**	**t**	**P> |t|**	**[0.025**	**0.975]**
Intercept	-0.0813	0.012	-6.706	0.000	-0.105	-0.057
sentiment_mean_top_influencer	0.1099	0.028	3.908	0.000	0.054	0.166

Table B.62. Quantile 0.25 — LCID

Dep. Variable:	Return	**Pseudo R-squared:**	0.1600
Model:	QuantReg	**Bandwidth:**	0.04251
Method:	Least Squares	**Sparsity:**	0.1281
Date:	Mon, 27 Oct 2025	**No. Observations:**	68
Time:	08:58:28	**Df Residuals:**	66
		Df Model:	1

	coef	**std err**	**t**	**P> |t|**	**[0.025**	**0.975]**
Intercept	-0.0746	0.011	-6.636	0.000	-0.097	-0.052
sentiment_mean_top_influencer	0.1428	0.028	5.177	0.000	0.088	0.198

Table B.63. Quantile 0.5 — LCID

Dep. Variable:	Return	**Pseudo R-squared:**	0.2134
Model:	QuantReg	**Bandwidth:**	0.04465
Method:	Least Squares	**Sparsity:**	0.1055
Date:	Mon, 27 Oct 2025	**No. Observations:**	68
Time:	08:58:28	**Df Residuals:**	66
		Df Model:	1

	coef	**std err**	**t**	**P> |t|**	**[0.025**	**0.975]**
Intercept	-0.0406	0.012	-3.342	0.001	-0.065	-0.016
sentiment_mean_top_influencer	0.1340	0.031	4.300	0.000	0.072	0.196

Table B.64. Quantile 0.75 — LCID

Dep. Variable:	Return	**Pseudo R-squared:**	0.2003
Model:	QuantReg	**Bandwidth:**	0.04001
Method:	Least Squares	**Sparsity:**	0.1123
Date:	Mon, 27 Oct 2025	**No. Observations:**	68
Time:	08:58:28	**Df Residuals:**	66
		Df Model:	1

	coef	**std err**	**t**	**P> |t|**	**[0.025**	**0.975]**
Intercept	-0.0069	0.013	-0.526	0.601	-0.033	0.019
sentiment_mean_top_influencer	0.0871	0.035	2.469	0.016	0.017	0.158

Table B.65. Quantile 0.9 — LCID

Dep. Variable:	Return	**Pseudo R-squared:**	0.2593
Model:	QuantReg	**Bandwidth:**	0.04245
Method:	Least Squares	**Sparsity:**	0.2178
Date:	Mon, 27 Oct 2025	**No. Observations:**	68
Time:	08:58:28	**Df Residuals:**	66
		Df Model:	1

	coef	std err	t	P> |t|	[0.025	0.975]
Intercept	0.0085	0.021	0.395	0.694	-0.034	0.051
sentiment_mean_top_influencer	0.1239	0.057	2.163	0.034	0.010	0.238

Table B.66. OLS Regression Results for: LCID

Dep. Variable:	Return	**R-squared:**	0.345
Model:	OLS	**Adj. R-squared:**	0.335
Method:	Least Squares	**F-statistic:**	34.75
Date:	Mon, 27 Oct 2025	**Prob (F-statistic):**	1.41e-07
Time:	08:58:28	**Log-Likelihood:**	131.95
No. Observations:	68	**AIC:**	-259.9
Df Residuals:	66	**BIC:**	-255.5
Df Model:	1		
Covariance Type:	nonrobust		

	coef	std err	t	P> |t|	[0.025	0.975]
Intercept	-0.0402	0.008	-4.941	0.000	-0.056	-0.024
sentiment_mean_top_influencer	0.1229	0.021	5.895	0.000	0.081	0.165

Omnibus:	0.408	**Durbin-Watson:**	2.313
Prob(Omnibus):	0.815	**Jarque-Bera (JB):**	0.566
Skew:	0.048	**Prob(JB):**	0.754
Kurtosis:	2.564	**Cond. No.**	5.43

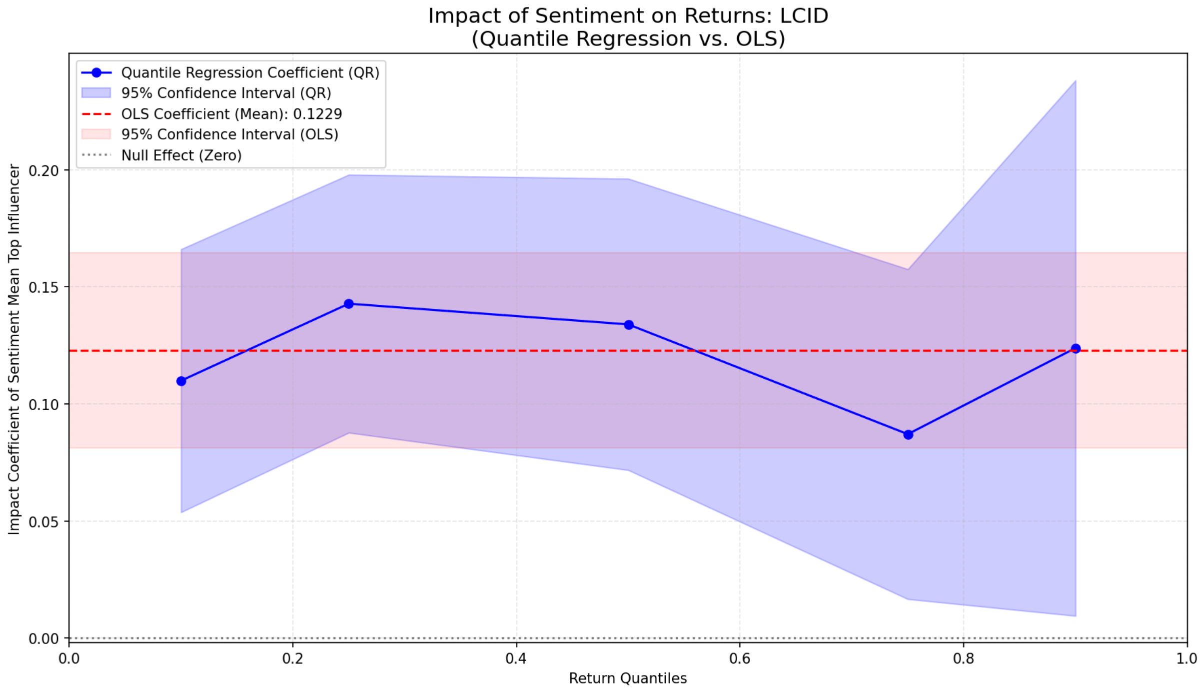

Figure B.11. Impact of Sentiment on Returns: LCID (Quantile Regression vs. OLS)

Table B.67. Quantile 0.1 — DIA

Dep. Variable:	Return	**Pseudo R-squared:**	0.2234
Model:	QuantReg	**Bandwidth:**	0.01709
Method:	Least Squares	**Sparsity:**	0.06850
Date:	Mon, 27 Oct 2025	**No. Observations:**	53
Time:	08:58:28	**Df Residuals:**	51
		Df Model:	1

	coef	**std err**	**t**	**P> |t|**	**[0.025**	**0.975]**
Intercept	0.0024	0.009	0.268	0.790	-0.015	0.020
sentiment_mean_top_influencer	0.0515	0.025	2.040	0.047	0.001	0.102

Table B.68. Quantile 0.25 — DIA

Dep. Variable:	Return	**Pseudo R-squared:**	0.1341
Model:	QuantReg	**Bandwidth:**	0.01317
Method:	Least Squares	**Sparsity:**	0.03818
Date:	Mon, 27 Oct 2025	**No. Observations:**	53
Time:	08:58:28	**Df Residuals:**	51
		Df Model:	1

	coef	**std err**	**t**	**P> |t|**	**[0.025**	**0.975]**
Intercept	0.0040	0.006	0.623	0.536	-0.009	0.017
sentiment_mean_top_influencer	0.0330	0.017	1.889	0.065	-0.002	0.068

Table B.69. Quantile 0.5 — DIA

Dep. Variable:	Return	**Pseudo R-squared:**	0.07983
Model:	QuantReg	**Bandwidth:**	0.01366
Method:	Least Squares	**Sparsity:**	0.02963
Date:	Mon, 27 Oct 2025	**No. Observations:**	53
Time:	08:58:28	**Df Residuals:**	51
		Df Model:	1

	coef	**std err**	**t**	**P> |t|**	**[0.025**	**0.975]**
Intercept	0.0113	0.005	2.372	0.021	0.002	0.021
sentiment_mean_top_influencer	0.0320	0.012	2.565	0.013	0.007	0.057

Table B.70. Quantile 0.75 — DIA

Dep. Variable:	Return	**Pseudo R-squared:**	0.05855
Model:	QuantReg	**Bandwidth:**	0.01140
Method:	Least Squares	**Sparsity:**	0.03062
Date:	Mon, 27 Oct 2025	**No. Observations:**	53
Time:	08:58:28	**Df Residuals:**	51
		Df Model:	1

	coef	**std err**	**t**	**P> |t|**	**[0.025**	**0.975]**
Intercept	0.0153	0.004	4.064	0.000	0.008	0.023
sentiment_mean_top_influencer	0.0277	0.010	2.897	0.006	0.008	0.047

TABLE B.71. Quantile 0.9 — DIA

Dep. Variable:	Return	**Pseudo R-squared:**	0.05764
Model:	QuantReg	**Bandwidth:**	0.01608
Method:	Least Squares	**Sparsity:**	0.05199
Date:	Mon, 27 Oct 2025	**No. Observations:**	53
Time:	08:58:28	**Df Residuals:**	51
		Df Model:	1

	coef	**std err**	**t**	**P>|t|**	**[0.025**	**0.975]**
Intercept	0.0261	0.004	6.574	0.000	0.018	0.034
sentiment_mean_top_influencer	0.0403	0.011	3.638	0.001	0.018	0.063

TABLE B.72. OLS Regression Results for: DIA

Dep. Variable:	Return	**R-squared:**	0.180
Model:	OLS	**Adj. R-squared:**	0.164
Method:	Least Squares	**F-statistic:**	11.21
Date:	Mon, 27 Oct 2025	**Prob (F-statistic):**	0.00153
Time:	08:58:28	**Log-Likelihood:**	143.61
No. Observations:	53	**AIC:**	-283.2
Df Residuals:	51	**BIC:**	-279.3
Df Model:	1		
Covariance Type:	nonrobust		

	coef	**std err**	**t**	**P>|t|**	**[0.025**	**0.975]**
Intercept	0.0165	0.005	3.118	0.003	0.006	0.027
sentiment_mean_top_influencer	0.0463	0.014	3.348	0.002	0.019	0.074

Omnibus:	33.585	**Durbin-Watson:**	2.108
Prob(Omnibus):	0.000	**Jarque-Bera (JB):**	157.901
Skew:	1.444	**Prob(JB):**	5.16e-35
Kurtosis:	10.948	**Cond. No.**	6.88

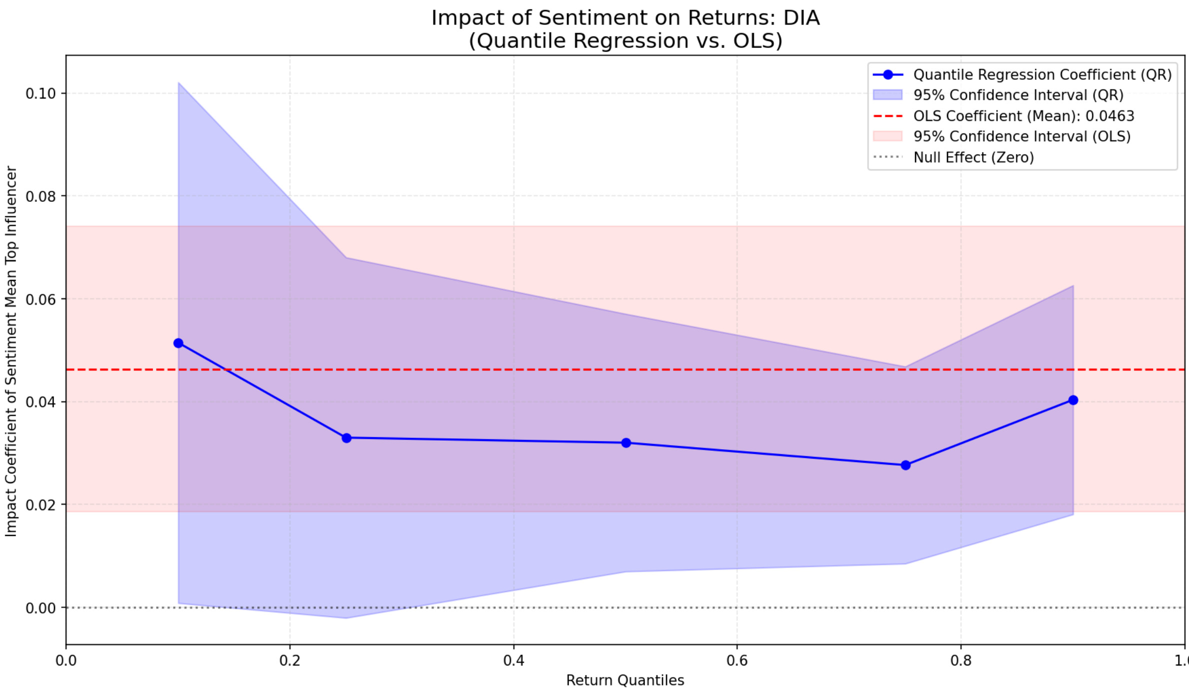

FIGURE B.12. Impact of Sentiment on Returns: DIA (Quantile Regression vs. OLS)

Table B.73. Quantile 0.1 — PLTR

Dep. Variable:	Return	**Pseudo R-squared:**	0.3753
Model:	QuantReg	**Bandwidth:**	nan
Method:	Least Squares	**Sparsity:**	nan
Date:	Mon, 27 Oct 2025	**No. Observations:**	33
Time:	08:58:28	**Df Residuals:**	31
		Df Model:	1

	coef	**std err**	**t**	**P> |t|**	**[0.025**	**0.975]**
Intercept	-0.0362	nan	nan	nan	nan	nan
sentiment_mean_top_influencer	0.0843	nan	nan	nan	nan	nan

Table B.74. Quantile 0.25 — PLTR

Dep. Variable:	Return	**Pseudo R-squared:**	0.3234
Model:	QuantReg	**Bandwidth:**	0.04896
Method:	Least Squares	**Sparsity:**	0.09886
Date:	Mon, 27 Oct 2025	**No. Observations:**	33
Time:	08:58:28	**Df Residuals:**	31
		Df Model:	1

	coef	**std err**	**t**	**P> |t|**	**[0.025**	**0.975]**
Intercept	-0.0153	0.008	-1.958	0.059	-0.031	0.001
sentiment_mean_top_influencer	0.1072	0.020	5.427	0.000	0.067	0.148

Table B.75. Quantile 0.5 — PLTR

Dep. Variable:	Return	**Pseudo R-squared:**	0.2765
Model:	QuantReg	**Bandwidth:**	0.05194
Method:	Least Squares	**Sparsity:**	0.09849
Date:	Mon, 27 Oct 2025	**No. Observations:**	33
Time:	08:58:28	**Df Residuals:**	31
		Df Model:	1

	coef	**std err**	**t**	**P> |t|**	**[0.025**	**0.975]**
Intercept	-0.0051	0.010	-0.507	0.616	-0.025	0.015
sentiment_mean_top_influencer	0.1154	0.036	3.178	0.003	0.041	0.189

Table B.76. Quantile 0.75 — PLTR

Dep. Variable:	Return	**Pseudo R-squared:**	0.2677
Model:	QuantReg	**Bandwidth:**	0.04897
Method:	Least Squares	**Sparsity:**	0.1178
Date:	Mon, 27 Oct 2025	**No. Observations:**	33
Time:	08:58:28	**Df Residuals:**	31
		Df Model:	1

	coef	**std err**	**t**	**P> |t|**	**[0.025**	**0.975]**
Intercept	0.0188	0.013	1.444	0.159	-0.008	0.045
sentiment_mean_top_influencer	0.1393	0.060	2.338	0.026	0.018	0.261

TABLE B.77. Quantile 0.9 — PLTR

Dep. Variable:	Return	**Pseudo R-squared:**	0.1770
Model:	QuantReg	**Bandwidth:**	nan
Method:	Least Squares	**Sparsity:**	nan
Date:	Mon, 27 Oct 2025	**No. Observations:**	33
Time:	08:58:28	**Df Residuals:**	31
		Df Model:	1

	coef	**std err**	**t**	**P> |t|**	**[0.025**	**0.975]**
Intercept	0.0309	nan	nan	nan	nan	nan
sentiment_mean_top_influencer	0.1514	nan	nan	nan	nan	nan

TABLE B.78. OLS Regression Results for: PLTR

Dep. Variable:	Return	**R-squared:**	0.451
Model:	OLS	**Adj. R-squared:**	0.434
Method:	Least Squares	**F-statistic:**	25.49
Date:	Mon, 27 Oct 2025	**Prob (F-statistic):**	1.87e-05
Time:	08:58:28	**Log-Likelihood:**	61.126
No. Observations:	33	**AIC:**	-118.3
Df Residuals:	31	**BIC:**	-115.3
Df Model:	1		
Covariance Type:	nonrobust		

	coef	**std err**	**t**	**P> |t|**	**[0.025**	**0.975]**
Intercept	-0.0021	0.008	-0.264	0.794	-0.018	0.014
sentiment_mean_top_influencer	0.1458	0.029	5.049	0.000	0.087	0.205

Omnibus:	24.504	**Durbin-Watson:**	1.479
Prob(Omnibus):	0.000	**Jarque-Bera (JB):**	48.337
Skew:	1.696	**Prob(JB):**	3.19e-11
Kurtosis:	7.863	**Cond. No.**	4.32

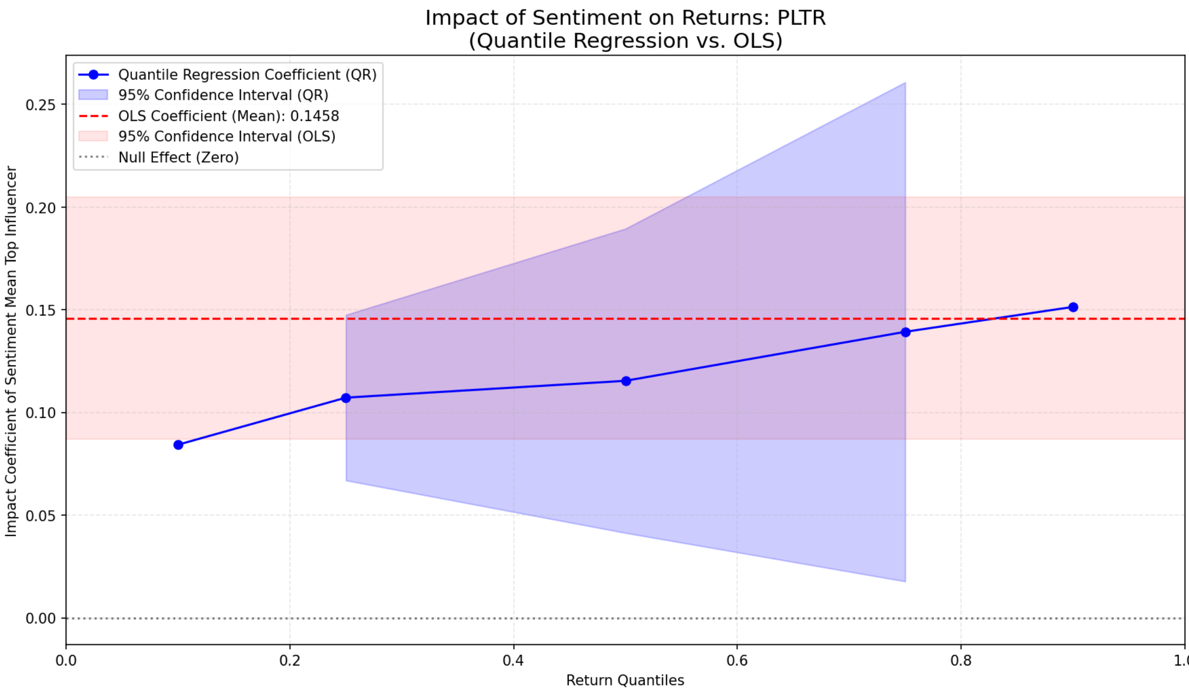

FIGURE B.13. Impact of Sentiment on Returns: PLTR (Quantile Regression vs. OLS)

Table B.79. Quantile 0.1 — UBER

Dep. Variable:	Return	**Pseudo R-squared:**	0.05269
Model:	QuantReg	**Bandwidth:**	0.01778
Method:	Least Squares	**Sparsity:**	0.1993
Date:	Mon, 27 Oct 2025	**No. Observations:**	97
Time:	08:58:29	**Df Residuals:**	95
		Df Model:	1

	coef	**std err**	**t**	**P> |t|**	**[0.025**	**0.975]**
Intercept	-0.0420	0.011	-3.990	0.000	-0.063	-0.021
sentiment_mean_top_influencer	0.0376	0.031	1.227	0.223	-0.023	0.098

Table B.80. Quantile 0.25 — UBER

Dep. Variable:	Return	**Pseudo R-squared:**	0.09167
Model:	QuantReg	**Bandwidth:**	0.01740
Method:	Least Squares	**Sparsity:**	0.07010
Date:	Mon, 27 Oct 2025	**No. Observations:**	97
Time:	08:58:29	**Df Residuals:**	95
		Df Model:	1

	coef	**std err**	**t**	**P> |t|**	**[0.025**	**0.975]**
Intercept	-0.0182	0.005	-3.429	0.001	-0.029	-0.008
sentiment_mean_top_influencer	0.0317	0.015	2.096	0.039	0.002	0.062

Table B.81. Quantile 0.5 — UBER

Dep. Variable:	Return	**Pseudo R-squared:**	0.1834
Model:	QuantReg	**Bandwidth:**	0.02091
Method:	Least Squares	**Sparsity:**	0.05126
Date:	Mon, 27 Oct 2025	**No. Observations:**	97
Time:	08:58:29	**Df Residuals:**	95
		Df Model:	1

	coef	**std err**	**t**	**P> |t|**	**[0.025**	**0.975]**
Intercept	-0.0119	0.004	-2.896	0.005	-0.020	-0.004
sentiment_mean_top_influencer	0.0559	0.012	4.800	0.000	0.033	0.079

Table B.82. Quantile 0.75 — UBER

Dep. Variable:	Return	**Pseudo R-squared:**	0.1546
Model:	QuantReg	**Bandwidth:**	0.01717
Method:	Least Squares	**Sparsity:**	0.05655
Date:	Mon, 27 Oct 2025	**No. Observations:**	97
Time:	08:58:29	**Df Residuals:**	95
		Df Model:	1

	coef	**std err**	**t**	**P> |t|**	**[0.025**	**0.975]**
Intercept	-0.0039	0.004	-0.954	0.342	-0.012	0.004
sentiment_mean_top_influencer	0.0673	0.012	5.680	0.000	0.044	0.091

Table B.83. Quantile 0.9 — UBER

Dep. Variable:	Return	**Pseudo R-squared:**	0.2179
Model:	QuantReg	**Bandwidth:**	0.02207
Method:	Least Squares	**Sparsity:**	0.1483
Date:	Mon, 27 Oct 2025	**No. Observations:**	97
Time:	08:58:29	**Df Residuals:**	95
		Df Model:	1

	coef	std err	t	P> |t|	[0.025	0.975]
Intercept	0.0078	0.009	0.889	0.376	-0.010	0.025
sentiment_mean_top_influencer	0.0967	0.028	3.496	0.001	0.042	0.152

Table B.84. OLS Regression Results for: UBER

Dep. Variable:	Return	**R-squared:**	0.256
Model:	OLS	**Adj. R-squared:**	0.248
Method:	Least Squares	**F-statistic:**	32.71
Date:	Mon, 27 Oct 2025	**Prob (F-statistic):**	1.23e-07
Time:	08:58:29	**Log-Likelihood:**	219.46
No. Observations:	97	**AIC:**	-434.9
Df Residuals:	95	**BIC:**	-429.8
Df Model:	1		
Covariance Type:	nonrobust		

	coef	std err	t	P> |t|	[0.025	0.975]
Intercept	-0.0154	0.004	-3.778	0.000	-0.023	-0.007
sentiment_mean_top_influencer	0.0661	0.012	5.720	0.000	0.043	0.089

Omnibus:	7.639	**Durbin-Watson:**	2.222
Prob(Omnibus):	0.022	**Jarque-Bera (JB):**	14.081
Skew:	0.142	**Prob(JB):**	0.000876
Kurtosis:	4.845	**Cond. No.**	4.82

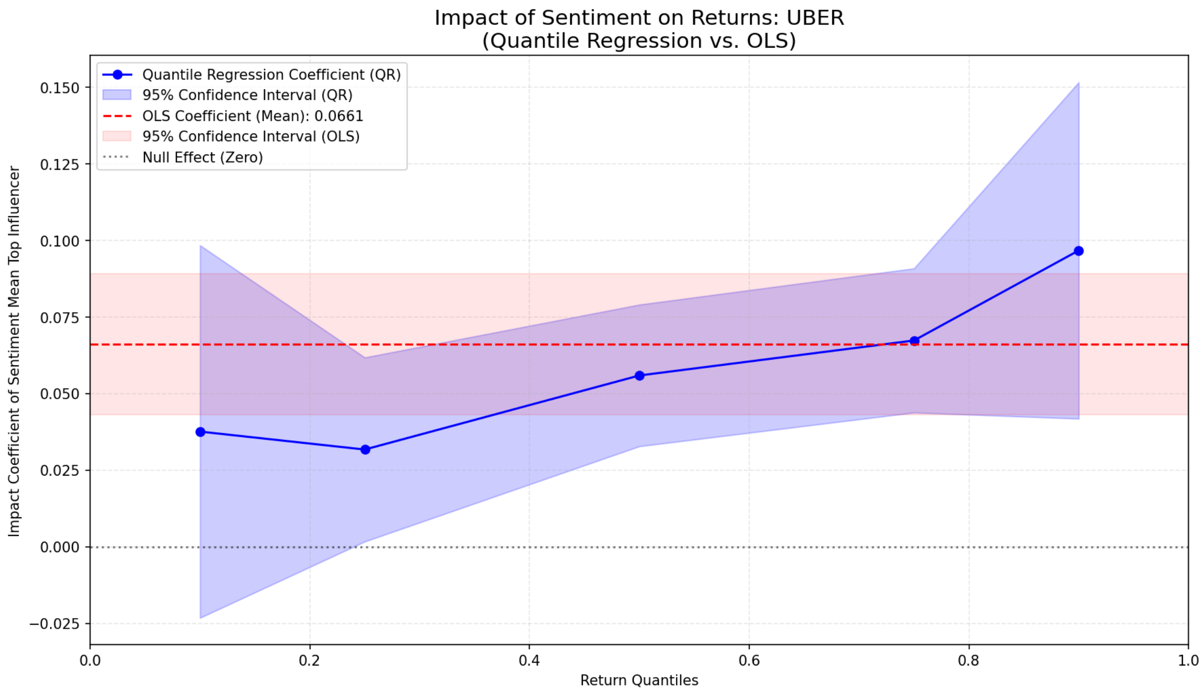

Figure B.14. Impact of Sentiment on Returns: UBER (Quantile Regression vs. OLS)

Table B.85. Quantile 0.1 — SOFI

Dep. Variable:	Return	**Pseudo R-squared:**	0.3820
Model:	QuantReg	**Bandwidth:**	0.04810
Method:	Least Squares	**Sparsity:**	0.1329
Date:	Mon, 27 Oct 2025	**No. Observations:**	52
Time:	08:58:34	**Df Residuals:**	50
		Df Model:	1

	coef	**std err**	**t**	**P>|t|**	**[0.025**	**0.975]**
Intercept	-0.0793	0.009	-8.896	0.000	-0.097	-0.061
sentiment_mean_top_influencer	0.1884	0.030	6.249	0.000	0.128	0.249

Table B.86. Quantile 0.25 — SOFI

Dep. Variable:	Return	**Pseudo R-squared:**	0.2820
Model:	QuantReg	**Bandwidth:**	0.03768
Method:	Least Squares	**Sparsity:**	0.09466
Date:	Mon, 27 Oct 2025	**No. Observations:**	52
Time:	08:58:34	**Df Residuals:**	50
		Df Model:	1

	coef	**std err**	**t**	**P>|t|**	**[0.025**	**0.975]**
Intercept	-0.0555	0.009	-6.478	0.000	-0.073	-0.038
sentiment_mean_top_influencer	0.1558	0.027	5.805	0.000	0.102	0.210

Table B.87. Quantile 0.5 — SOFI

Dep. Variable:	Return	**Pseudo R-squared:**	0.2279
Model:	QuantReg	**Bandwidth:**	0.03392
Method:	Least Squares	**Sparsity:**	0.07538
Date:	Mon, 27 Oct 2025	**No. Observations:**	52
Time:	08:58:34	**Df Residuals:**	50
		Df Model:	1

	coef	**std err**	**t**	**P>|t|**	**[0.025**	**0.975]**
Intercept	-0.0197	0.008	-2.400	0.020	-0.036	-0.003
sentiment_mean_top_influencer	0.1026	0.026	3.907	0.000	0.050	0.155

Table B.88. Quantile 0.75 — SOFI

Dep. Variable:	Return	**Pseudo R-squared:**	0.1678
Model:	QuantReg	**Bandwidth:**	0.03137
Method:	Least Squares	**Sparsity:**	0.08066
Date:	Mon, 27 Oct 2025	**No. Observations:**	52
Time:	08:58:34	**Df Residuals:**	50
		Df Model:	1

	coef	**std err**	**t**	**P>|t|**	**[0.025**	**0.975]**
Intercept	-0.0038	0.008	-0.481	0.632	-0.020	0.012
sentiment_mean_top_influencer	0.0860	0.026	3.350	0.002	0.034	0.138

TABLE B.89. Quantile 0.9 — SOFI

Dep. Variable:	Return	**Pseudo R-squared:**	0.1065
Model:	QuantReg	**Bandwidth:**	0.04299
Method:	Least Squares	**Sparsity:**	0.1306
Date:	Mon, 27 Oct 2025	**No. Observations:**	52
Time:	08:58:34	**Df Residuals:**	50
		Df Model:	1

	coef	**std err**	**t**	**P> |t|**	**[0.025**	**0.975]**
Intercept	0.0235	0.009	2.503	0.016	0.005	0.042
sentiment_mean_top_influencer	0.0549	0.031	1.752	0.086	-0.008	0.118

TABLE B.90. OLS Regression Results for: SOFI

Dep. Variable:	Return	**R-squared:**	0.273
Model:	OLS	**Adj. R-squared:**	0.258
Method:	Least Squares	**F-statistic:**	18.77
Date:	Mon, 27 Oct 2025	**Prob (F-statistic):**	7.10e-05
Time:	08:58:34	**Log-Likelihood:**	93.537
No. Observations:	52	**AIC:**	-183.1
Df Residuals:	50	**BIC:**	-179.2
Df Model:	1		
Covariance Type:	nonrobust		

	coef	**std err**	**t**	**P> |t|**	**[0.025**	**0.975]**
Intercept	-0.0264	0.009	-2.968	0.005	-0.044	-0.009
sentiment_mean_top_influencer	0.1233	0.028	4.333	0.000	0.066	0.181

Omnibus:	42.541	**Durbin-Watson:**	2.228
Prob(Omnibus):	0.000	**Jarque-Bera (JB):**	258.689
Skew:	1.879	**Prob(JB):**	6.70e-57
Kurtosis:	13.260	**Cond. No.**	5.33

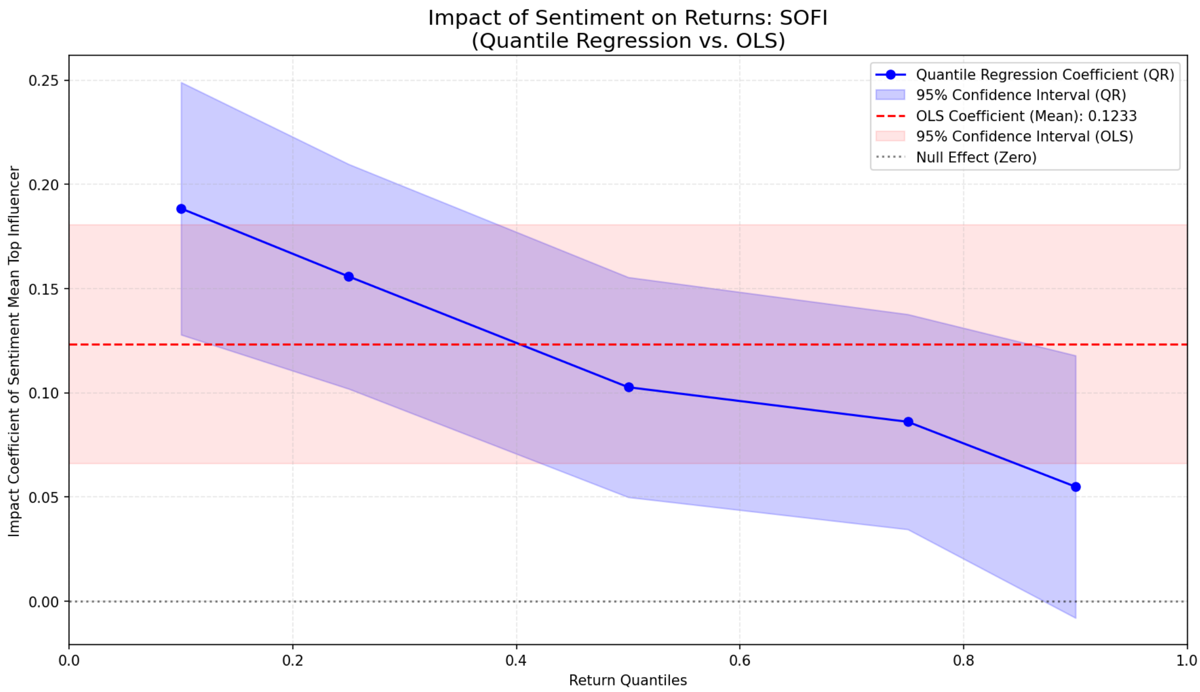

FIGURE B.15. Impact of Sentiment on Returns: SOFI (Quantile Regression vs. OLS)

Table B.91. Quantile 0.1 — NKE

Dep. Variable:	Return	**Pseudo R-squared:**	0.2237
Model:	QuantReg	**Bandwidth:**	0.03502
Method:	Least Squares	**Sparsity:**	0.09595
Date:	Mon, 27 Oct 2025	**No. Observations:**	60
Time:	08:58:35	**Df Residuals:**	58
		Df Model:	1

	coef	**std err**	**t**	**P> |t|**	**[0.025**	**0.975]**
Intercept	-0.0425	0.005	-9.405	0.000	-0.052	-0.033
sentiment_mean_top_influencer	0.1060	0.017	6.244	0.000	0.072	0.140

Table B.92. Quantile 0.25 — NKE

Dep. Variable:	Return	**Pseudo R-squared:**	0.1447
Model:	QuantReg	**Bandwidth:**	0.02840
Method:	Least Squares	**Sparsity:**	0.07893
Date:	Mon, 27 Oct 2025	**No. Observations:**	60
Time:	08:58:35	**Df Residuals:**	58
		Df Model:	1

	coef	**std err**	**t**	**P> |t|**	**[0.025**	**0.975]**
Intercept	-0.0308	0.005	-6.028	0.000	-0.041	-0.021
sentiment_mean_top_influencer	0.0746	0.018	4.246	0.000	0.039	0.110

Table B.93. Quantile 0.5 — NKE

Dep. Variable:	Return	**Pseudo R-squared:**	0.1080
Model:	QuantReg	**Bandwidth:**	0.02879
Method:	Least Squares	**Sparsity:**	0.06270
Date:	Mon, 27 Oct 2025	**No. Observations:**	60
Time:	08:58:35	**Df Residuals:**	58
		Df Model:	1

	coef	**std err**	**t**	**P> |t|**	**[0.025**	**0.975]**
Intercept	-0.0094	0.005	-1.929	0.059	-0.019	0.000
sentiment_mean_top_influencer	0.0546	0.019	2.951	0.005	0.018	0.092

Table B.94. Quantile 0.75 — NKE

Dep. Variable:	Return	**Pseudo R-squared:**	0.1546
Model:	QuantReg	**Bandwidth:**	0.02622
Method:	Least Squares	**Sparsity:**	0.07028
Date:	Mon, 27 Oct 2025	**No. Observations:**	60
Time:	08:58:35	**Df Residuals:**	58
		Df Model:	1

	coef	**std err**	**t**	**P> |t|**	**[0.025**	**0.975]**
Intercept	0.0032	0.005	0.595	0.554	-0.007	0.014
sentiment_mean_top_influencer	0.0538	0.022	2.408	0.019	0.009	0.099

Table B.95. Quantile 0.9 — NKE

Dep. Variable:	Return	**Pseudo R-squared:**	0.1712
Model:	QuantReg	**Bandwidth:**	0.03109
Method:	Least Squares	**Sparsity:**	0.1241
Date:	Mon, 27 Oct 2025	**No. Observations:**	60
Time:	08:58:35	**Df Residuals:**	58
		Df Model:	1

	coef	std err	t	P> |t|	[0.025	0.975]
Intercept	0.0175	0.008	2.314	0.024	0.002	0.033
sentiment_mean_top_influencer	0.0674	0.033	2.033	0.047	0.001	0.134

Table B.96. OLS Regression Results for: NKE

Dep. Variable:	Return	**R-squared:**	0.243
Model:	OLS	**Adj. R-squared:**	0.230
Method:	Least Squares	**F-statistic:**	18.63
Date:	Mon, 27 Oct 2025	**Prob (F-statistic):**	6.26e-05
Time:	08:58:35	**Log-Likelihood:**	125.56
No. Observations:	60	**AIC:**	-247.1
Df Residuals:	58	**BIC:**	-242.9
Df Model:	1		
Covariance Type:	nonrobust		

	coef	std err	t	P> |t|	[0.025	0.975]
Intercept	-0.0138	0.005	-2.924	0.005	-0.023	-0.004
sentiment_mean_top_influencer	0.0774	0.018	4.317	0.000	0.041	0.113

Omnibus:	17.143	**Durbin-Watson:**	2.784
Prob(Omnibus):	0.000	**Jarque-Bera (JB):**	75.445
Skew:	-0.398	**Prob(JB):**	4.14e-17
Kurtosis:	8.436	**Cond. No.**	4.68

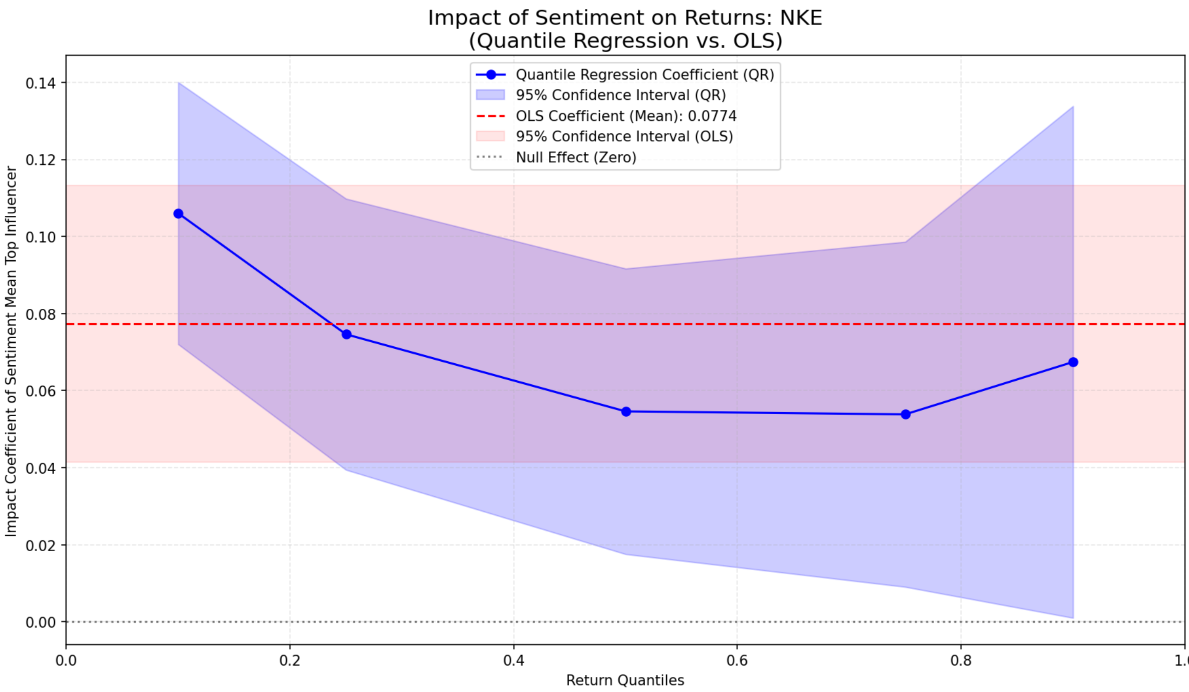

Figure B.16. Impact of Sentiment on Returns: NKE (Quantile Regression vs. OLS)

Table B.97. Quantile 0.1 — PFE

Dep. Variable:	Return	**Pseudo R-squared:**	0.2265
Model:	QuantReg	**Bandwidth:**	0.02003
Method:	Least Squares	**Sparsity:**	0.08241
Date:	Mon, 27 Oct 2025	**No. Observations:**	73
Time:	08:58:35	**Df Residuals:**	71
		Df Model:	1

	coef	**std err**	**t**	**P> |t|**	**[0.025**	**0.975]**
Intercept	-0.0195	0.003	-6.363	0.000	-0.026	-0.013
sentiment_mean_top_influencer	0.0896	0.028	3.237	0.002	0.034	0.145

Table B.98. Quantile 0.25 — PFE

Dep. Variable:	Return	**Pseudo R-squared:**	0.1634
Model:	QuantReg	**Bandwidth:**	0.01794
Method:	Least Squares	**Sparsity:**	0.05603
Date:	Mon, 27 Oct 2025	**No. Observations:**	73
Time:	08:58:35	**Df Residuals:**	71
		Df Model:	1

	coef	**std err**	**t**	**P> |t|**	**[0.025**	**0.975]**
Intercept	-0.0110	0.003	-3.779	0.000	-0.017	-0.005
sentiment_mean_top_influencer	0.0649	0.021	3.027	0.003	0.022	0.108

Table B.99. Quantile 0.5 — PFE

Dep. Variable:	Return	**Pseudo R-squared:**	0.1450
Model:	QuantReg	**Bandwidth:**	0.01740
Method:	Least Squares	**Sparsity:**	0.04159
Date:	Mon, 27 Oct 2025	**No. Observations:**	73
Time:	08:58:35	**Df Residuals:**	71
		Df Model:	1

	coef	**std err**	**t**	**P> |t|**	**[0.025**	**0.975]**
Intercept	0.0009	0.002	0.380	0.705	-0.004	0.006
sentiment_mean_top_influencer	0.0556	0.017	3.275	0.002	0.022	0.089

Table B.100. Quantile 0.75 — PFE

Dep. Variable:	Return	**Pseudo R-squared:**	0.06520
Model:	QuantReg	**Bandwidth:**	0.01481
Method:	Least Squares	**Sparsity:**	0.04571
Date:	Mon, 27 Oct 2025	**No. Observations:**	73
Time:	08:58:35	**Df Residuals:**	71
		Df Model:	1

	coef	**std err**	**t**	**P> |t|**	**[0.025**	**0.975]**
Intercept	0.0092	0.002	3.962	0.000	0.005	0.014
sentiment_mean_top_influencer	0.0457	0.014	3.381	0.001	0.019	0.073

TABLE B.101. Quantile 0.9 — PFE

Dep. Variable:	Return	**Pseudo R-squared:**	0.04480
Model:	QuantReg	**Bandwidth:**	0.01736
Method:	Least Squares	**Sparsity:**	0.08041
Date:	Mon, 27 Oct 2025	**No. Observations:**	73
Time:	08:58:35	**Df Residuals:**	71
		Df Model:	1

	coef	**std err**	**t**	**P>|t|**	**[0.025**	**0.975]**
Intercept	0.0194	0.003	6.585	0.000	0.014	0.025
sentiment_mean_top_influencer	0.0372	0.017	2.160	0.034	0.003	0.072

TABLE B.102. OLS Regression Results for: PFE

Dep. Variable:	Return	**R-squared:**	0.217
Model:	OLS	**Adj. R-squared:**	0.206
Method:	Least Squares	**F-statistic:**	19.63
Date:	Mon, 27 Oct 2025	**Prob (F-statistic):**	3.34e-05
Time:	08:58:35	**Log-Likelihood:**	197.94
No. Observations:	73	**AIC:**	-391.9
Df Residuals:	71	**BIC:**	-387.3
Df Model:	1		
Covariance Type:	nonrobust		

	coef	**std err**	**t**	**P>|t|**	**[0.025**	**0.975]**
Intercept	0.0002	0.002	0.088	0.930	-0.004	0.004
sentiment_mean_top_influencer	0.0589	0.013	4.431	0.000	0.032	0.085

Omnibus:	3.487	**Durbin-Watson:**	1.883
Prob(Omnibus):	0.175	**Jarque-Bera (JB):**	2.873
Skew:	-0.285	**Prob(JB):**	0.238
Kurtosis:	3.786	**Cond. No.**	6.97

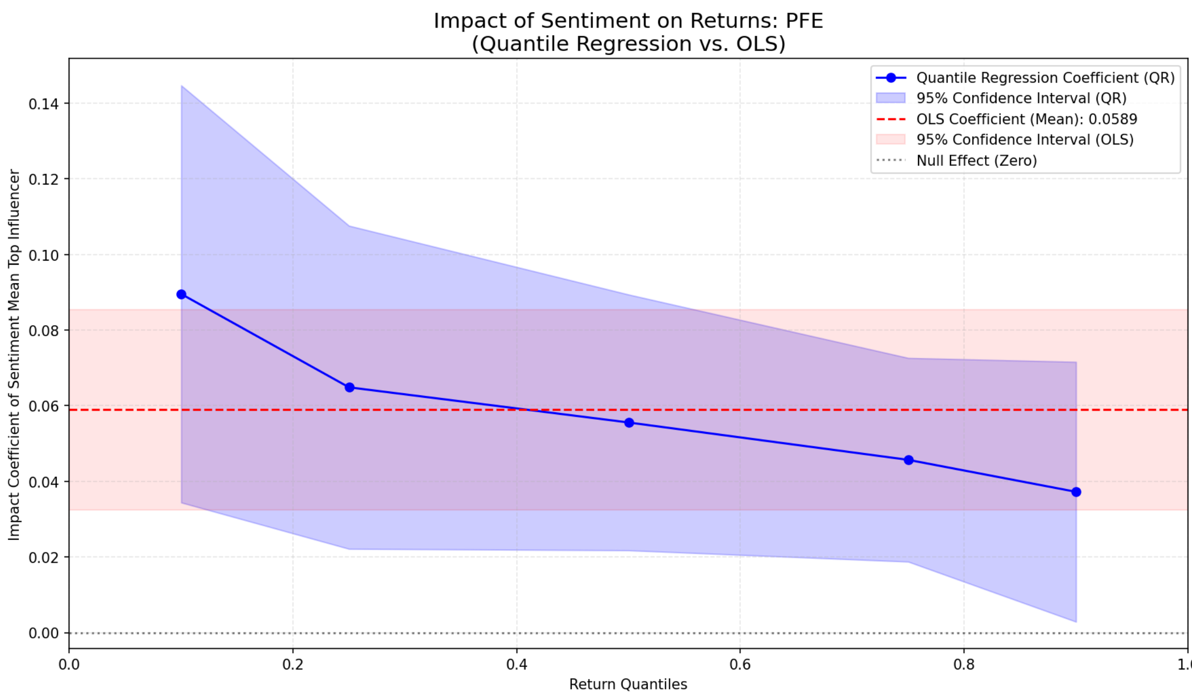

FIGURE B.17. Impact of Sentiment on Returns: PFE (Quantile Regression vs. OLS)