Data Science in Finance and Economics

https://www.aimspress.com/journal/dsfe

DSFE, 5(4): 577–600.
DOI:10.3934/DSFE.2025023
Received: 22 April 2025
Revised: 17 August 2025
Accepted: 25 November 2025
Published: 22 December 2025

Research article

Integrating financial sentiment analysis with coreference resolution: A comprehensive empirical framework

Ruicheng Liu and Erik Cambria*

College of Computing and Data Science, Nanyang Technological University, 50 Nanyang Ave, Block N 4, Singapore

* **Correspondence:** Email: cambria@ntu.edu.sg.

Abstract: Financial sentiment analysis has emerged as a crucial tool for deciphering the complex narratives within financial markets. However, the accuracy of sentiment attribution to specific financial entities remains a significant challenge. This study introduces a novel framework that integrates state-of-the-art coreference resolution (CR) with a robust causal inference methodology—propensity score matching (PSM) and difference-in-differences (DID)—to precisely measure the impact of entity-specific sentiment. Coreference resolution is employed to accurately attribute sentiments to financial entities, significantly enhancing the precision of the sentiment signal. By leveraging a financial domain-adapted language model (FinBERT) and a comprehensive analytical framework, we demonstrate the causal impact of CR-enhanced sentiment on key financial metrics. To address the "black-box" nature of predictive models, we further employ SHapley Additive exPlanations (SHAP) to interpret our Random Forest sentiment predictions, identifying the primary drivers of market sentiment. The findings contribute to a deeper understanding of sentiment-market dynamics, revealing the nuanced, causal impact of sentiment on market behavior. The implications of this study extend beyond academic contributions, offering valuable, interpretable insights for investment strategies, risk management, and regulatory oversight. This research represents a significant step forward in applying advanced, data-driven, and interpretable techniques to financial decision-making.

Keywords: financial sentiment analysis; propensity score matching; difference-in-differences analysis; natural language processing; language models; financial markets

JEL Codes: C55, G14, G41, C21

1. Introduction

This study explores the application of coreference resolution (CR) in financial sentiment analysis. We integrate CR as a critical preprocessing step to improve entity-level sentiment attribution in financial texts such as earnings reports and news articles. By resolving pronouns and ambiguous references before sentiment scoring, we aim to enhance the accuracy and reliability of the extracted financial signal, which is then used within a causal inference framework to assess its market impact.

In the rapidly advancing financial industry, the deluge of data ranging from intricate transactional records to the vast sprawl of news articles and social media discourse has introduced significant challenges and opportunities. The critical task of extracting actionable insights from such voluminous and complex datasets has become increasingly indispensable for stakeholders in the financial ecosystem (Gupta et al., 2020, Xing et al., 2018). Sentiment analysis, particularly, stands out as a vital analytical tool in decoding the multifaceted narratives within financial markets, enabling a deeper comprehension of market sentiments and investor behaviors: Using diverse data sources, including financial reports, news feeds, and social media platforms to gauge the prevailing market mood, thereby assisting investors, financial institutions, and regulators in making well-informed decisions. This analytical approach not only facilitates the prediction of market trends, but also aids in monitoring potential market manipulations or fraudulent activities, underpinning a more transparent and efficient financial environment.

The field of financial sentiment analysis has seen a remarkable transformation with the introduction of advanced natural language processing (NLP) and deep learning techniques. Initially grounded in lexicon-based methods, the advent of machine learning models marked the beginning of a new era. The development of pre-trained language models, such as BERT (Devlin et al., 2019) and its subsequent adaptations like FinBERT (Liu et al., 2020), has significantly improved the accuracy of sentiment analysis. These models, trained on vast amounts of text, can understand the context and nuances of language, offering insights that were previously unattainable.

However, while these contributions are pivotal, they highlight persistent gaps in the literature. Previous studies have often struggled with two key issues: (1) accurately attributing document-level sentiment to the specific entities mentioned within, a problem that sophisticated CR can solve; and (2) moving beyond correlation to establish a causal link between sentiment and market outcomes. The evolving landscape of market conditions necessitates methodologies that not only precisely identify entities but also rigorously control for the confounding factors inherent in observational financial data. Our research aims to fill this gap by proposing a sophisticated analytical framework that transcends traditional sentiment analysis approaches.

The main contributions of this work are as follows:

1. **Methodological Innovation** We introduce a novel framework that combines state-of-the-art CR (the Link-Append system) with a robust causal inference methodology (PSM-DID). This integration allows for the precise measurement of the impact of entity-specific sentiment by mitigating selection bias and controlling for unobserved heterogeneity, moving the analysis from correlation to causation.
2. **Empirical Findings** Our comprehensive analysis reveals the causal pathways and dynamic nature of sentiment's effect on financial metrics. This is supported by mediation analysis, which dissects the channels of influence, and spatial models, which test for sentiment spillovers across economically linked firms.

3. **Enhanced Model Interpretability** We address the "black-box" nature of sentiment prediction models by employing SHapley Additive exPlanations (SHAP). This new contribution provides interpretable insights into the key drivers of financial sentiment, offering actionable intelligence for market participants beyond mere predictive accuracy.
4. **Robustness and Generalizability** We conduct a suite of rigorous robustness checks, including alternative model specifications and time frames, to confirm the stability and reliability of our findings. This ensures that the observed effects are not artifacts of a particular model or sample.

This endeavor not only broadens the scope of financial sentiment analysis but also enhances the precision, causality, and interpretability of sentiment attribution in the context of rapidly changing financial narratives.

2. Related works

2.1. Coreference resolution

The intersection of CR and financial sentiment analysis has been a burgeoning area of interest within the computational linguistics and financial analytics communities. Early strides in CR, as established by Clark and Manning (2016), laid the groundwork for subsequent innovations in natural language processing (NLP) technologies. These foundational models were integrated into practical tools like the spaCy library by Press and Wolf (2017). However, the financial domain, with its specialized lexicon and high degree of ambiguity, presents unique challenges that demand more nuanced approaches. The OntoNotes (Hovy et al., 2006) project provided a critical benchmark dataset for testing CR models across various genres, including financial texts.

Recent advancements have been driven by deep learning architectures. Models like ELMo (Peters et al., 2018) and BERT (Devlin et al., 2019) significantly pushed the boundaries of CR accuracy, with subsequent models like SpanBERT (Joshi et al., 2020) further enhancing performance by focusing on span representation. More recent innovations have reframed CR as a question-answering task (Wu et al., 2020) or introduced more computationally efficient word-level models (Dobrovolskii, 2021).

A significant breakthrough, and the one adopted in this study, was achieved by Bohnet et al. (2023), who introduced a text-to-text (seq2seq) transition-based system using the T5 language model. This approach, which jointly predicts mentions and links them into coreference clusters, achieved new state-of-the-art accuracy on benchmark datasets and has shown impressive performance in zero-shot and few-shot settings. As these technologies evolve, they enable more sophisticated applications in domains like finance, where accurately linking a pronoun like "it" or "the company" to the correct corporate entity is critical for valid sentiment analysis.

2.2. Financial sentiment analysis

In the evolving field of financial sentiment analysis, the integration of sophisticated NLP techniques has become standard (Du et al., 2025). The development of domain-specific models like FinBERT (Liu et al., 2020) exemplifies this trend, leveraging deep learning to decode the complexities of financial language. Researchers have applied advanced models to tasks ranging from fraud detection (Jiang et al., 2019) to market prediction (Zerveas et al., 2021). Despite these strides, the role of CR in enhancing sentiment analysis accuracy has remained a fertile area for research, aiming to fine-tune the precision of NLP applications in finance.

However, a key limitation of many existing studies is their reliance on correlational analysis. To establish a more robust, causal understanding of sentiment's impact, quasi-experimental methods are required. Propensity score matching (PSM) is a statistical technique used to estimate the effect of a treatment (e.g., exposure to negative news sentiment) by accounting for covariates that predict receiving the treatment (Rosenbaum and Rubin, 1983). In finance, PSM can be used to match companies with similar characteristics (e.g., size, industry) but differing sentiment exposure, thereby creating a more valid control group. The difference-in-differences (DID) method further strengthens causal claims by comparing the change in outcomes over time between a treated group and a control group. This approach controls for time-invariant unobserved factors that might influence company performance, helping to isolate the true effect of the sentiment shock.

Our study is situated at the confluence of these fields. We argue that while advancements in CR and sentiment analysis are powerful, their full potential is only unlocked when integrated into a rigorous causal inference framework like PSM-DID. This combination allows us to move beyond simply measuring sentiment to quantifying its causal effect on financial outcomes, a critical step for generating truly actionable insights.

3. Methodology

3.1. Overall research design

To isolate the causal effect of accurately attributed financial sentiment on market outcomes, this study employs a multi-stage, quasi-experimental research design. The framework first leverages a state-of-the-art CR system to enhance sentiment attribution, then uses a Propensity score matching (PSM) and difference-in-differences (DID) approach to mitigate selection bias and control for unobserved heterogeneity. The process is structured as follows: (1) Pre-process a large corpus of financial texts using a high-precision CR model to resolve entity references. (2) Calculate entity-specific sentiment scores using a domain-adapted language model. (3) Employ a PSM-DID framework to estimate the causal impact of sentiment on financial performance metrics. (4) Conduct extensive robustness checks and supplementary analyses, including mediation and spatial models, to validate the findings. (5) Use SHAP to interpret the machine learning models, adding a layer of explainability to the predictions.

3.2. Coreference resolution pre-processing

We begin by assembling an extensive dataset of financial news articles and company filings from the Factiva database, focusing on companies within the S&P 500 index. After standard text cleaning (e.g., tokenization, stop word removal), we apply a state-of-the-art CR model to accurately link all mentions (e.g., "Apple," "the company," "it") to the correct underlying entity.

For this task, we adopt the Link-Append system proposed by Bohnet et al. (2023), which represents the cutting edge in seq2seq transition-based CR. We chose this model for two primary reasons: its state-of-the-art accuracy and its high efficiency in processing long documents, which is essential for our corpus of news articles. The system processes a document one sentence at a time, taking the current sentence and the previously identified coreference clusters as input, and then generating a sequence of actions to update the clusters. The action space consists of three core operations:

- Link Action: Connects a mention in the current sentence to a previous mention in the text, effectively creating a new link in a coreference chain. For example, it might link "the tech giant" in sentence five to “Apple Inc.” in sentence one.

- Append Action: Adds a mention from the current sentence to an already established coreference cluster. For example, if a cluster for “Apple Inc.” already exists, a mention of “it” would be appended to this cluster.
- SHIFT Action: Signals that the processing for the current sentence is complete. The system then “shifts” to the next sentence, carrying forward the updated cluster information.

This incremental, sentence-by-sentence approach eliminates the need for a separate mention-detection step and efficiently handles the long-form text common in financial news, ensuring that sentiment is attributed with the highest possible precision.

3.3. Entity-Specific sentiment analysis and causal inference framework

Leveraging the enhanced CR output, we perform sentiment analysis using FinBERT, a financial sector-specific adaptation of the BERT model. This step involves calculating sentiment scores for each entity within a given timeframe, based on the contextual analysis of sentences where the company is mentioned directly or identified through CR. The sentiment score for an entity e in a document d can be computed as:

$$S(e,d) = \frac{1}{N}\sum_{i=1}^{N} s(x_i) \tag{1}$$

where $S(e,d)$ represents the overall sentiment score for entity e in document d, N is the number of mentions of entity e in document d, and $s(x_i)$ is the sentiment score of the i-th mention of entity e, obtained from FinBERT.

To move from correlation to causation, we employ a robust empirical framework. The core challenge is that news coverage and its associated sentiment are not random. To address this, we first use Propensity score matching (PSM) (Rosenbaum and Rubin, 1983). The propensity score, or the conditional probability of receiving the treatment (e.g., a high-sentiment news event) given observed covariates, is estimated using a logistic regression model:

$$e(X) = Pr(D = 1|X) = \frac{1}{1 + e^{-\beta X}} \tag{2}$$

where D is the binary treatment indicator, X is the vector of firm-level covariates (e.g., size, industry, past performance), and β is the vector of coefficients. By matching treated firms with control firms that have a similar propensity score, we create a balanced comparison group, mitigating selection bias.

On the other hand, the difference-in-differences (DID) approach is a quasi-experimental technique used to estimate the effect of a specific intervention or treatment by comparing the changes in outcomes over time between a treated group and a control group (Card and Krueger, 1994). The basic formula for DID estimation is:

$$\delta = (\bar{Y}^t_{post} - \bar{Y}^t_{pre}) - (\bar{Y}^c_{post} - \bar{Y}^c_{pre}) \tag{3}$$

where δ is the DID estimator, $\bar{Y}^t_{post}$ and $\bar{Y}^t_{pre}$ are the average outcomes for the treated group in the post-treatment and pre-treatment periods, respectively, and $\bar{Y}^c_{post}$ and $\bar{Y}^c_{pre}$ are the average outcomes for the control group in the post-treatment and pre-treatment periods, respectively. DID controls for time-invariant unobserved heterogeneity, which is crucial in financial sentiment analysis, as it helps to isolate the causal impact of sentiments on company performance by accounting for any pre-existing differences between the treated and control groups.

3.4. Robustness and supplementary analyses

To ensure the validity of our findings, we conduct a series of additional analyses.

Mediation Analysis: We employ mediation effect regression to investigate the mechanisms through which sentiment influences financial outcomes. This involves estimating a system of equations to determine if the effect of sentiment (X) on a financial outcome (Y) operates through an intermediary variable, or mediator (M). The key equations are:

$$Y = \alpha_1 + \beta_1 X + \epsilon_1 \tag{4}$$
$$M = \alpha_2 + \beta_2 X + \epsilon_2 \tag{5}$$
$$Y = \alpha_3 + \beta_3 X + \gamma M + \epsilon_3 \tag{6}$$

where Y is the outcome variable (financial performance), X is the independent variable (sentiment), M is the mediator variable, α_i are the intercepts, β_i and γ are the coefficients, and ϵ_i are the error terms.

Equation (4) represents the total effect of sentiment (X) on financial outcomes (Y). Equation (5) represents the effect of sentiment (X) on the mediator variable (M). Equation (6) represents the effect of sentiment (X) and the mediator (M) on financial outcomes (Y), while controlling for the direct effect of sentiment.

The mediation effect, also known as the indirect effect, is quantified by the product of coefficients β_2 and γ. It measures the extent to which the mediator variable (M) accounts for the relationship between sentiment (X) and financial outcomes (Y). A significant mediation effect suggests that sentiment influences financial outcomes through the mediator variable.

Dynamic and Spatial Models: To explore the persistence of sentiment states, we use a Markov transition probability matrix, which calculates the probability of transitioning from one sentiment state (e.g., positive) to another over time. The research extends to examining the Markov transition probability matrix and spatial Durbin model results post-CR, offering novel insights into the dynamic nature of financial sentiment and its spatial correlations. The Markov transition probability matrix, denoted as P, is defined as:

$$P = \begin{bmatrix} p_{11} & p_{12} & \cdots & p_{1n} \\ p_{21} & p_{22} & \cdots & p_{2n} \\ \vdots & \vdots & \ddots & \vdots \\ p_{n1} & p_{n2} & \cdots & p_{nn} \end{bmatrix}, \tag{7}$$

where p_{ij} represents the probability of transitioning from state i to state j, and $\sum_{j=1}^{n} p_{ij} = 1$ for all i.

Spatial Durbin model, which incorporates spatial lag and spatial error terms, can be expressed as:

$$y = \rho W y + X\beta + WX\theta + \epsilon \tag{8}$$

where y is the dependent variable, X is the matrix of independent variables, W is the spatial weight matrix, ρ is the spatial autoregressive coefficient, β and θ are the coefficient vectors, and ϵ is the error term.

These supplementary analyses, combined with multivariate regression and other robustness checks, provide a comprehensive and validated picture of the role of sentiment in financial markets.

4. Financial datasets

4.1. Factiva database for financial news analysis

The Factiva database, produced by Dow Jones, is an essential resource for financial analysis. With access to over 30,000 sources worldwide, including news articles, company filings, financial data, and industry reports from more than 200 countries in 30+ languages, Factiva provides a comprehensive view of global financial markets. Its powerful search engine and advanced filtering options streamline information retrieval, while its data export capabilities in CSV, Excel, and XML formats make it invaluable for researchers and analysts incorporating financial news data into their work. Factiva's extensive coverage, sophisticated search features, and flexible data export options make it a crucial tool for anyone requiring up-to-date financial information and insights.

4.2. Text data processing

To construct a comprehensive corpus of finance-related articles, we gathered raw newspaper articles from the Factiva database spanning the period from 2020 to 2022. This data collection effort focused on articles directly linked to companies listed within the S&P 500 index, ensuring the financial relevance of the corpus. To further refine the data sources, we restricted our selection to articles published by prominent financial news providers, such as the Wall Street Journal, the New York Times, USA Today, and the Washington Post. Each article was meticulously linked to at least one corresponding company, and some articles were associated with multiple S&P 500 companies due to their broader coverage of industry-wide developments. This initial data collection process yielded a substantial dataset comprising 200,642 article-company pairs.

To enhance the quality and relevance of the dataset for subsequent financial analysis, we implemented a rigorous data cleaning process. We conducted a manual review of the most frequently occurring headlines to identify and remove articles that deviated from the desired focus on finance-related information. This eliminated opinion pieces, product reviews, and stock market summaries, ensuring the dataset's alignment with the objectives of our study. Furthermore, we employed the Harvard General Inquirer2 word lists "Legal" and "@Econ" to systematically count the frequency of legal and finance-related terms within each article. Articles with less than 5% of their words classified as legal or finance-related were excluded, ensuring that the corpus remained centered on information pertinent to investors. Additionally, adhering to the recommendations of Tetlock et al. (2008) and Loughran and McDonald (2016), we removed articles containing fewer than 50 words, as textual analysis of extremely short texts often yields unreliable results. These meticulous filtering steps resulted in a refined dataset comprising 150,374 article-company pairs, providing a robust foundation for our subsequent analyses.

Finally, we applied standard text-cleaning methods commonly utilized in natural language processing and finance to prepare the article bodies for further analysis. These methods involved the systematic removal of URLs, email addresses, numbers, punctuation, special symbols, and common English stop words. Additionally, we standardized all words to lowercase, replaced multiple white spaces with a single space, expanded contractions, and removed possessives ("'s") and hyphens. These preprocessing steps ensured that the text data was consistent, structured, and free from extraneous elements, thereby enhancing the effectiveness of our subsequent computational analyses.

4.3. Data summary

To compile the necessary data for our analysis, we obtained S&P 500 constituents, stock returns, and trading volumes from the CRSP database. Fundamental data and accounting information were retrieved from Compustat, while Fama & French factors (Fama and French, 1993) and risk-free rates were sourced from Kenneth French's website. Additionally, we utilized Harvard General Inquirer's "Econ@" and "Legal" lists to identify economics, business, and legal terms, respectively. For sentiment analysis, we employed Loughran & McDonald's negative and positive word lists Loughran and McDonald (2016). To count company mentions and their locations within articles, we used a search vector composed of company names and their variations, as identified from CRSP common names and our newspaper article database.

Assessing the impact of our CR method—specifically, its efficacy in identifying instances where a text segment pertains to a particular company—on sentiment metrics necessitates the availability of company-level sentiment estimates within the same article. As discussed previously, a single article may encompass discussions on multiple companies with varying tones. Consequently, obtaining an unbiased estimation of company-level tone becomes imperative to gauge the extent to which CR enhances the alternative approach of uniformly applying the same tone to all companies featured in a given article.

To ensure clarity, all variables used in the analysis are defined in Table 1.

Table 1. *Variable Definitions*

Variable Name	Description	Unit
AbnormalReturn	Market Model Residual.	%
MarketSentiment	A measure of bull market sentiment.	%
FinancialPositiveIndex	A composite index of positive financial indicators.	%
FiscalNegativeIndex	A composite index of negative fiscal indicators.	%
EconomicBenefit	Economic benefits derived from the 'FinancialPositiveIndex'.	$
TextCorpusNumeric	A numerical value derived from the financial text corpus.	%
CorefFactor	A factor representing the influence of CR.	%
CorpusTimeSpan	The time span value of the financial corpus.	Day

Table 2. Descriptive statistics for the analyzed variables.

Variable Name	Mean	Std. Dev.	Min	Max	Skewness	Kurtosis
AbnormalReturn	1.179	0.636	0.006	9.662	1.055	5.271
MarketSentiment	0.140	0.267	0.000	1.080	1.696	4.147
FinancialPositiveIndex	43.613	14.851	3.836	84.850	0.246	2.160
FiscalNegativeIndex	1.309	0.601	0.667	7.912	3.016	15.112
EconomicBenefit	10.030	1.625	10.465	18.119	1.161	6.371
TextCorpusNumeric	0.356	0.168	0.018	1.197	0.192	3.157
CorefFactor	6.073	7.823	0.075	60.203	−1.069	19.698
CorpusTimeSpan	7.808	7.348	0.000	21.280	0.457	1.934

Table 2 presents the descriptive statistics for these key variables. The statistics reveal a diverse set of distributions, with several variables exhibiting significant skewness and kurtosis, a common feature in financial data reflecting "fat-tailed" distributions where extreme events are more common. Our subsequent regression analyses employ robust standard errors to account for this characteristic.

5. Results

Our analysis commenced with the collection of an extensive dataset of news headlines directly related to individual companies within the S&P 500, covering the period from January 1, 2020, to December 31, 2022. The initial dataset comprised 418,703 headlines, which, upon undergoing our refined entity recognition process, expanded to include 576,191 instances of enriched news texts. These texts were meticulously analyzed for sentiment extraction, employing a FinBERT model specifically tuned for the financial domain. This model, selected for its superior accuracy in financial sentiment analysis, facilitated a comprehensive evaluation of sentiment across the dataset, allowing us to accurately gauge the prevailing market sentiments associated with these entities over the specified period.

Table 3. PSM-DID pre-matching balance test.

Variable	Group	Mean	Mean Difference	T-value	P-value
EconomicBenefit	Treatment	14.591	−0.709	−4.704	0.021
	Control	13.882			
FiscalNegativeIndex	Treatment	1.291	0.138	3.612	0.049
	Control	1.153			

Table 3 presents a PSM-DID pre-matching balance test comparing experimental and control groups across two variables, EconomicBenefit and FiscalNegativeIndex. For EconomicBenefit, despite the experimental group having a higher mean (14.591) compared to the control group (13.882), resulting in a mean difference of −4.704, the T-value (0.537) and P-value (0.676) indicate that this difference is not statistically significant. In the case of FiscalNegativeIndex, the experimental group's mean is slightly higher (1.291) than that of the control group (1.153), with a mean difference of 0.612. However, the extremely high P-value (1.234) alongside a negligible T-value (0.039) suggests that this observed difference lacks statistical significance, questioning the impact of the intervention. This analysis demonstrates that, at least on the surface, the interventions may not have had a significant effect on the measured outcomes, as indicated by the high P-values, suggesting the need for further investigation or reconsideration of the intervention's efficacy.

Table 4. PSM-DID Post-matching balance test.

Variable Name	Group	Mean	Mean Difference	T-value	P-value
EconomicBenefit	Experimental	12.583	−5.320	0.462	0.670
	Control	11.052			
FiscalNegativeIndex	Experimental	1.876	0.474	0.038	0.684
	Control	1.243			

The results from a PSM-DID post-matching balance test are summarized in Table 4, aimed at evaluating the comparability of experimental and control groups across variables EconomicBenefit and FiscalNegativeIndex. This test is crucial for causal inference studies, ensuring that observed differences post-treatment are attributable to the intervention. For EconomicBenefit, the experimental group exhibits a mean of 12.583 compared to the control group's 11.052, with a mean difference of −5.320, a T-value of 0.462, and a P-value of 0.670. These statistics indicate no significant difference between groups, highlighting an effective matching process. Similarly, FiscalNegativeIndex shows a mean of 1.876 for the experimental group and 1.243 for the control, with a mean difference of 0.474, a T-value of 0.038, and a P-value of 0.684, reinforcing the matching process's success. The non-significant P-values for both variables (> 0.05) confirm the efficacy of the PSM-DID method in creating statistically comparable groups, laying a solid foundation for attributing observed outcome differences to the treatment effect with minimal confounding from pre-treatment disparities.

Table 5 presents the results of a PSM-DID baseline regression analysis, aimed at evaluating the impact of an intervention across different time frames. The coefficients for SE*time, significant across all models, suggest a varying effect of the intervention over time. The significance levels, denoted by asterisks, underscore the robustness of these findings. The SE coefficient shows variability in the intervention's impact, indicating its effects become more pronounced and consistent over narrower time frames. The Time variable shows a negative but only sometimes significant relationship, hinting at a potential overall downward trend. The inclusion of Year and Industry as control variables ensures that the analysis accounts for fixed effects. The constant term is highly significant across all models, indicating a substantial effect not explained by the included variables. The sample size and R^2 values provide insights into the robustness and explanatory power of the regression.

Table 5. PSM-DID baseline regression.

	(1)	(2)	(3)
	[−36,36]	[−24,24]	[−18,18]
MarketSentiment * Time	−0.154***	−0.082**	−0.070**
	(−3.425)	(−2.021)	(−1.819)
MarketSentiment	0.226**	0.337**	0.424**
	(2.040)	(2.081)	(1.852)
Time	−0.031	−0.025**	−0.026***
	(−1.200)	(−1.912)	(−2.149)
Constant	0.560***	0.571***	0.788***
	(6.300)	(9.198)	(7.634)
Year Fixed Effects	Y	Y	Y
Industry Fixed Effects	Y	Y	Y
Observations	20,440	20,828	19,264
R-squared	0.110	0.130	0.168

Note: T-statistics are in parentheses. *** p<0.01, ** p<0.05, * p<0.1.

In the presented analysis of Table 6, employing a PSM-DID fixed effects model, seven distinct specifications were systematically explored to assess the influence of various predictors on the dependent variable. Each model specification consistently incorporated the AbnormalReturn variable, alongside the interaction term SE×time, revealing a statistically significant negative impact across all configurations. This finding signifies a consistent diminishment in the outcome variable attributable to the interaction between time and standard error, with significance levels rigorously denoted as $p < 0.10$, $p < 0.05$, and $p < 0.01$ for *, **, and ***, respectively. Furthermore, the SE variable itself exhibited a positive and significant influence across all models, indicating that an increase in the standard error correlates with an elevation in the dependent variable. The inclusion of additional variables such as CorefFactor, FinancialPositiveIndex, Age, Fiscal Negative Index, EconomicBenefit, and TextCorpusNumeric, each introduced in individual model specifications, exerted varying degrees of influence on the dependent variable, thereby underscoring their predictive power. Notably, the significant positive effect of Fiscal Negative Index and the negative impact of EconomicBenefit in specific models highlight the nuanced interplay between these variables and the outcome of interest. The incorporation of Year and Industry as fixed effects served to control for unobservable heterogeneity, thereby enhancing the reliability of the estimations. Despite the consistent sample size ($N = 17253$) and an R^2 value of 0.053 across all models indicating a degree of explained variance, a considerable portion of the outcome variability remains unexplained, pointing towards the complex and multifaceted nature of the dependent variable's determinants. This analysis underscores the critical importance of considering both the temporal dynamics and the intrinsic variability within the data to elucidate the underlying mechanisms influencing the outcome variable, thereby contributing to a more nuanced understanding of the subject matter within the scientific discourse.

Table 6. PSM-DID fixed effects model.

	(1)	(2)	(3)	(4)	(5)	(6)	(7)
	AbnormalReturn	AbnormalReturn	AbnormalReturn	AbnormalReturn	AbnormalReturn	AbnormalReturn	AbnormalReturn
SE*time	−0.235**	−0.215*	−0.232**	−0.246**	−0.234**	−0.232**	−0.210**
	(−2.72)	(−2.16)	(−2.91)	(−2.34)	(−2.73)	(−2.54)	(−2.62)
SE	0.173**	0.146*	0.176**	0.126*	0.135**	0.146**	0.164**
	(2.00)	(1.91)	(2.01)	(1.96)	(2.05)	(2.01)	(1.96)
time	−0.0870***	−0.0574***	−0.0920***	−0.0550***	−0.0622***	−0.0922***	−0.0784***
	(−3.42)	(−3.34)	(−3.40)	(−3.34)	(−3.91)	(−3.40)	(−3.63)
CorefFactor	0.00143						0.00252
	(0.87)						(1.23)
FinancialPositiveIndex		−0.00714**					−0.00564**
		(−2.35)					(−2.46)
Age			−0.0421				−0.0151
			(−0.21)				(−0.21)
FiscalNegativeIndex				0.136***			0.142***
				(7.51)			(8.00)
EconomicBenefit					−0.164***		−0.143***
					(−4.27)		(−3.30)
TextCorpusNumeric						−0.000520	0.269**
						(−0.00)	(2.25)
Year	Y	Y	Y	Y	Y	Y	Y
Industry	Y	Y	Y	Y	Y	Y	Y
Cons	0.879***	1.228***	0.820**	0.654***	2.864***	0.901***	2.460***
	(4.23)	(5.28)	(2.35)	(3.49)	(6.23)	(5.41)	(3.82)
N	17253	17253	17253	17253	17253	17253	17253
R^2	0.053	0.053	0.053	0.053	0.053	0.053	0.053

Note: T-statistics are in parentheses. *** p<0.01, ** p<0.05, * p<0.1.

Table 7 offers a comprehensive examination of the mediation effects within a regression model framework, across five distinct specifications. This analysis aims to uncover the nuanced roles played by variables such as AbnormalReturn, AMI, and Indhold in influencing the model's outcomes. Notably, the presence of significant predictors is confirmed through their coefficients and statistical significance levels, with asterisks ($p < 0.01$) marking a high degree of significance. Particularly in models (1), (3), and (5), AbnormalReturn emerges as a central variable, consistently demonstrating a negative and statistically significant impact on the dependent variable. This pattern highlights AbnormalReturn's critical mediation role, with coefficients like −0.0306*** in model (1) and −0.0516*** in model (5) reinforcing the variable's substantial influence across various contexts.

Table 7. Mediation effect regression model.

	(1) AbnormalReturn	(2) AMI	(3) AbnormalReturn	(4) Indhold	(5) AbnormalReturn
On	−0.0306***	−0.0124***	−0.0478***	1.4743***	−0.0516***
	(−3.04)	(−2.93)	(−3.41)	(−7.09)	(−3.71)
CorefFactor	0.00532***	−0.00472***	0.00218***	0.25092***	0.00438***
	(−5.12)	(−13.28)	(−2.41)	(−12.92)	(−3.75)
FinancialPositiveIndex	0.00159***	0.00614***	0.0049***	0.15855***	0.00188***
	(−1.81)	(−18.97)	(−5.05)	(−13.53)	(−1.98)
Age	0.0207***	−0.0457***	0.0104***	3.2331***	0.0232***
	(−10.76)	(−34.91)	(−3.54)	(−49.26)	(−8.28)
FiscalNegativeIndex	0.134***	−0.169***	0.089***	7.452***	0.127***
EconomicBenefit	−0.0734***	−0.4153***	−0.2386***	2.794***	−0.0928***
	(−7.43)	(−83.05)	(−21.61)	(−11.39)	(−5.77)
TextCorpusNumeric	0.499***	0.457***	0.682***	2.125***	0.349***
AMI	(−8.22)	(−18.65)		(−2.05	(−7.01)
				(−11.58)	
Indhold			−0.424		−0.000273
			(−34.80)		(−0.94)
Year	Y	Y	Y	Y	Y
Industry	Y	Y	Y	Y	Y
Cons	1.686***	−3.969***	0.00401	−55.69***	1.670***
	(−17.35)	(−65.64)	(−0.03)	(−23.33)	(−16.35)
N	32017	32017	32017	32017	32017
R^2	0.242	0.172	0.462	0.328	0.196

Note: T-statistics are in parentheses. *** p<0.01, ** p<0.05, * p<0.1.

Conversely, AMI, featured in models (2) and (5), displays a diverse impact. Model (2) reveals a significant negative effect (−0.0124***), suggesting AMI's capacity to significantly alter the mediation pathway. In stark contrast, Indhold, predominantly included in model (4), exhibits a strong positive effect (1.4743***), indicating a significant mediating role distinct from those observed with AbnormalReturn and AMI. The analysis further delves into the effects of other pivotal variables like CorefFactor, FinancialPositiveIndex, Age, FiscalNegativeIndex, and EconomicBenefit TextCorpusNumeric, each uniquely contributing to the mediation effect across the models. For instance, the variables Age and FiscalNegativeIndex show significant effects in both directions, underscoring the complexity of their roles in the mediation process. The inclusion of Year and Industry as fixed effects across all models serves to mitigate potential confounding influences, thereby bolstering the reliability of the findings. The variation in constants across the models, from 1.686*** in model (1) to −55.69*** in model (4), reflects the distinct baseline levels inherent to each model's context. With a consistent sample size (N=32017) and R² values ranging from 0.172 in model (2) to 0.462 in model (3), the analysis highlights the models' differential explanatory capacities and the intricate relationships at play. These insights afford a deeper comprehension of the mediation effects exerted by various variables, enriching our understanding of the mechanisms underpinning the observed outcomes.

Table 8. Robustness check.

	−1	−2	−3
	full sample	[−36,36]	[−18,18]
SE*time	−0.126***	−0.235***	−0.274*
	(−2.55)	(−2.14)	(−2.29)
SE	0.066**	0.169	0.141*
time	(−2.59)	(−1.89)	(−2.01)
CorefFactor	−0.0363**	−0.0316	−0.1220***
	(−1.78)	(−1.79)	(−3.47)
FinancialPositiveIndex	0.00294***	0.00013	0.00231
	(−5.92)	(−0.06)	(−1.14)
Age	0.00190***	0.00146	-0.00606**
	(−3.51)	(−1.23)	(−1.71)
FiscalNegativeIndex	0.0228	-0.0911	-0.0110
	(-1.08)	(-2.59)	(-0.22)
EconomicBenefit	0.108***	0.168***	0.176***
	(−23.40)	(−15.32)	(−6.10)
TextCorpusNumeric	−0.00929	−0.01150	−0.15498**
	(−0.89)	(−0.59)	(−3.16)
	0.0801***	−0.0762	0.3146**
	(−3.66)	(−0.81)	(−2.78)
Year	Y	Y	Y
Industry	Y	Y	Y
Cons	1.862***	1.357***	3.168***
	(−15.47)	(−6.36)	(−5.5)
N	56523	27664	14333
R^2	0.1	0.1	0.1

Note: T-statistics are in parentheses. *** p<0.01, ** p<0.05, * p<0.1.

The robustness check outlined in Table 8 evaluates the stability of the effects across different sample sizes and time frames, specifically focusing on the full sample, a mid-range period ($[-36, 36]$), and a short-range period ($[-18, 18]$). The SE×time interaction demonstrates a statistically significant negative impact across all time frames, with the effect size increasing as the window narrows, indicating a stronger interaction effect in more immediate proximity to the event ($-0.274*$ in the $[-18, 18]$ period). The SE variable shows variability in its significance across different periods, highlighting its fluctuating influence.

Notably, the variable EconomicBenefit consistently exhibits a significant positive effect across all models, underscoring its robustness ($0.108***$, $0.168***$, and $0.176***$), whereas other variables like CorefFactor and FinancialPositiveIndex Age display mixed results in terms of significance, suggesting a more complex relationship with the dependent variable. The inclusion of Year and Industry as fixed effects across all specifications ensures control for potential confounding factors, enhancing the credibility of the findings. The constants across models ($1.862***$, $1.357***$, $3.168***$) further affirm the presence of a baseline effect that remains significant regardless of the time frame considered. The sample sizes and R^2 values (0.1 across all models) indicate a moderate explanatory power, with the models capturing a consistent portion of the variability in the dependent variable across varying contexts. This robustness check thus validates the stability and reliability of the observed effects, contributing to a nuanced understanding of the underlying dynamics at play.

Table 9. Post-Financial sentiment analysis DID regression.

	(1) AbnormalReturn	(2) AbnormalReturn	(3) AbnormalReturn	(4) AbnormalReturn	(5) AbnormalReturn	(6) AbnormalReturn	(7) AbnormalReturn
SE*time	-0.180**	-0.281**	-0.174**	-0.225**	-0.286**	-0.296**	-0.176**
	(-2.47)	(-2.30)	(-1.78)	(-2.49)	(-1.65)	(-1.84)	(-2.29)
SE	0.175**	0.185**	0.242**	0.167**	0.270**	0.187**	0.159**
	(1.862)	(1.751)	(2.288)	(2.688)	(2.576)	(2.441)	(1.883)
time	-0.038**	-0.036**	-0.042**	-0.048**	-0.062**	-0.037**	-0.054**
	(-2.66)	(-2.43)	(-1.67)	(-2.03)	(-2.30)	(-1.71)	(-2.40)
CorefFactor	0.00142						0.00161
	(0.76)						(0.71)
FinancialPositiveIndex		-0.0057**					-0.00457*
		(-2.36)					(-1.72)
Age			-0.0251				-0.00347
			(-0.21)				(-0.07)
FiscalNegativeIndex				0.136***			0.162***
				(8.96)			(7.41)
EconomicBenefit					-0.165***		-0.163***
					(-4.25)		(-3.16)
TextCorpusNumeric						-0.173	0.176
						(-1.45)	(1.31)
Year	Y	Y	Y	Y	Y	Y	Y
Industry	Y	Y	Y	Y	Y	Y	Y
Cons	0.942***	1.143***	0.894**	0.725***	2.718***	0.269***	2.314***
	(5.95)	(4.75)	(4.50)	(5.80)	(4.60)	(5.80)	(5.75)
N	17413	17413	17413	17413	17413	17413	17413
R²	0.042	0.042	0.042	0.042	0.042	0.042	0.042

Note: T-statistics are in parentheses. *** p<0.01, ** p<0.05, * p<0.1.

The analysis encapsulated in Table 9 showcases a DID regression focusing on the post-financial sentiment across seven distinct models, all under the umbrella of the AbnormalReturn1 variable. The SE×time interaction across these models consistently demonstrates a statistically significant negative impact, with coefficients ranging from -0.180** to -0.296**, highlighting the temporal and conditional variability in sentiment post-financial events. This effect is underscored by the consistent significance of the SE variable itself, indicating a robust relationship between standard errors and the outcome variable across all models.

Further analysis reveals nuanced relationships with other variables across the models. For instance, CorefFactor and FinancialPositiveIndex exhibit conditional significance, indicating varying influences on the outcome based on the model. Specifically, FinancialPositiveIndex shows a negative effect in models (2) and (7), suggesting sector-specific temporal dynamics. Additionally, FiscalNegativeIndex and EconomicBenefit variables emerge with strong positive and negative impacts respectively, indicating their crucial roles in driving sentiment post-financial events. Notably, the constant terms across models signify the baseline sentiment level, with significant variability indicating the diverse contexts of the financial sentiment analysis. The uniformity in sample size (N=17413) and R^2 values (0.042) across all models suggests a consistent explanatory power, albeit modest, underscoring the complexity of financial sentiment dynamics as captured by the DID regression analysis.

Table 10. Markov Transition Probability Matrix after Coreference Resolution.

Spacial Lags	t \t+1	n	1	2	3	4
1	1	678	0.6254	0.3054	0.0062	0.1524
	2	421	0.2015	0.5214	0.0048	0.3625
	3	201	0.0023	0.3145	0.1425	0.4955
	4	65	0.01425	0.0241	0.1552	0.5162
2	1	152	0.5218	0.6254	0.2565	0.6029
	2	431	0.0062	0.4429	0.1144	0.7824
	3	405	0.0047	0.5932	0.9524	0.8145
	4	152	0.0001	0.6041	0.1152	0.9933
3	1	15	0.5026	0.2014	0.2004	0.4582
	2	99	0.0954	0.3026	0.0036	0.9246
	3	272	0.0262	0.0426	0.0042	0.8556
	4	130	0.0002	0.4814	0.0142	0.2651
4	1	6	0.2059	0.5241	0.6254	0.4157
	2	42	0.0841	0.8012	0.7152	0.3311
	3	182	0.0847	0.6231	0.4821	0.2514
	4	569	0.0042	0.5471	0.2695	0.3369

The Markov transition probability matrix, showcased in Table 10, elucidates the intricacies of spatial entity transitions across different temporal phases, structured around four spatial lags. This analytical framework is instrumental in quantifying the probabilities of transitioning from one state at time t to another at time $t + 1$. It is pivotal for deciphering the underlying mechanisms of spatial dynamics, highlighting the tendencies of states to either maintain their current status or evolve into new states, with these probabilities significantly influenced by spatial proximity.

The matrix reveals that certain transitions have notably high probabilities, indicating a pronounced tendency for entities to either remain in their current state or move to a particular future state. For example, a significant transition probability is observed at spatial lag 2, where the transition probability from state 1 to state 4 is notably high (0.6029), suggesting a considerable likelihood of advancement or change to a distinct state over time. Conversely, lower transition probabilities across various segments imply a degree of unpredictability in state evolution, underscoring the complex interplay between spatial dynamics and external influences. The disparities in sample sizes across spatial lags and states introduce an additional layer of complexity, where smaller sample sizes may result in less reliable estimates of transition probabilities. Thus, this matrix serves as an essential analytical tool for comprehending the temporal shifts of spatial entities, offering invaluable insights for predictive modeling and strategic decision-making in spatial research.

Table 11. Results of the spatial durbin model after Coreference Resolution. control variables and their spatial terms, time fixed effects and individual fixed effects are controlled. SWM stands for spatial weight matrix, wald test 1 is SAR with P-value, wald test 2 is SEM with P-value, wald test 3 is joint parameter test with P-value. S is sample size.

	Baseline Regression				Lagged Explanatory Variables			
	Model1	Model2	Model3	Model4	Model5	Model6	Model7	Model8
SWM	W_d^H	W_e	W_{ge}^H	W_w^H	W_w^H	W_d^H	W_{ge}^H	W_w^H
P	0.3625^b	0.2265^a	0.1834^a	0.2985^a	0.2063^a	0.4825^a	0.5985^a	0.555^a
	(0.2013)	(0.0452)	(0.0655)	(0.0958)	(0.0459)	(0.0365)	(0.5985)	(0.2795)
IS	−0.3485	-0.2552^b	−0.01625	0.0265^c	-0.0624^a	-0.7695^a	-0.3662^a	-0.3555^b
	(0.12254)	(0.0013)	(0.0123)	(0.04529)	(0.2458)	(0.1365)	(0.2995)	(0.20588)
STQ	-0.8625^a	-0.7825^a	-0.8495^a	-0.7065^a	-0.2849^a	-0.2265^a	-0.4825^a	-0.5825^b
	(0.2036)	(0.09546)	(0.1452)	(0.2044)	(0.2245)	(0.1595)	(0.4965)	(0.6953)
OS	−0.02365	−0.04265	0.02658	−0.04562	−0.0652a	−0.0455a	−0.7598a	0.3775a
	(0.0265)	(0.0958)	(0.02652)	(0.01459)	(0.0265)	(0.2758)	(0.4458)	(0.4443)
ES	−0.3695a	−0.2256a	−0.2695a	−0.2499a	0.4595	−0.5956b	−0.3555a	−0.3625a
	(0.0426)	(0.0365)	(0.0425)	(0.004295)	(0.01552)	(0.04855)	(0.03253)	(0.2575)
UR	−0.4692a	−0.4985a	−0.6042a	−0.6025a	−0.2265a	0.0425a	−0.8525a	−1.325a
	(0.2215)	(0.1125)	(0.1409)	(0.2154)	(0.2066)	(0.04295)	(0.2559)	(0.5554)
UR2	−0.4692b	0.04925a	0.5985a	0.08845a	0.0625a	−0.9588a	−0.5525a	−0.7955a
	(0.0263)	(0.01459)	(0.01958)	(0.04655)	(0.01254)	(0.00452)	(0.2558)	(0.2544)
TI	0.0642b	0.04595a	0.04265a	0.03451a	0.04255b	−0.3795a	−0.355a	−0.5725b
	(0.0045)	(0.00495)	(0.00695)	(0.00485)	(0.00785)	(0.2985)	(0.0065)	(0.553)
EE	0.2954a	0.2236a	0.2451a	0.3695a	0.2366a	−1.355a	−1.3625a	0.5525a
	(0.04252)	(0.01692)	(0.0499)	(0.0152)	(0.00459)	(0.2253)	(0.2458)	(0.255)
wIS	-0.4026^a	-0.2036^a	-0.0954^a	-0.0955^b	-0.9355^a	0.2958^a	-0.5525^a	-0.5525^b
	(0.06245)	(0.02384)	(0.01295)	(0.04525)	(0.0655)	(0.2582)	(0.2355)	(0.2525)
wKL	0.5369	1.2685^a	0.8695^a	0.8954^a	1.2655^b	-0.6951^a	-0.3625^a	-0.5525^a
	(0.2013)	(0.0185)	(0.2236)	(0.1952)	(0.3495)	(0.01256)	(0.2953)	(0.2555)
wOS	0.3625^b	−0.03655	−0.0459	-0.0452^a	-0.0452^a	-0.1225^a	-0.955^a	0.59625^a
	(0.4406)	(0.06254)	(0.0485)	(0.2955)	(0.0652)	(0.4265)	(0.5553)	(0.2783)
wES	−0.6629	-0.2694^a	-0.3475^a	-0.2265^a	0.3625^b	-0.7925^a	0.3575^a	-0.5955^a
	(0.2695)	(0.06245)	(0.0654)	(0.0625)	(0.2013)	(0.1254)	(0.365)	(0.4983)
wUR	2.0362^b	1.6245^a	0.7062^a	1.9522^a	-0.6958^a	-0.3672^a	-0.7825^a	-0.7625^a
	(0.8953)	(0.2635)	(0.3062)	(0.4035)	(0.2558)	(0.2655)	(0.2693)	(0.2055)
wUR2	-0.3795^b	-0.1425^a	-0.2065^a	-0.2369^a	-0.3625^a	0.4855^a	-0.4485^a	-0.3625^c
	(0.2265)	(0.04655)	(0.2013)	(0.0065)	(0.2758)	(0.14993)	(0.8713)	(0.2553)
wTI	0.1694^a	0.02654	0.0695	0.0425^b	0.1275^a	0.79855^a	-0.3795^a	-0.9525^a
	(0.0492)	(0.0125)	(0.00452)	(0.0205)	(0.04525)	(0.2142)	(0.20455)	(0.2555)
wEE	0.4395^b	−0.3495	-0.37649	0.00299	0.37955^b	-0.7825^a	-0.3445^a	-1.2225^a
	(0.1895)	(0.03652)	(0.2013)	(0.0896)	(0.1245)	(0.9853)	(0.2785)	(0.2958)
Const	0.03625^a	0.3798^a	0.04552^a	0.0362^a	0.04525^a	-0.3565^a	0.47285^a	-0.5525^a
	(0.00125)	(0.00203)	(0.00425)	(0.004253)	(0.00123)	(0.2485)	(0.4283)	(0.2583)
Wald Test 1	69.23	162.45	143.26	156.42	61.62	136.42	121.42	120.44
	(0.000)	(0.000)	(0.000)	(0.002)	(0.000)	(0.002)	(0.000)	(0.000)
Wald Test 2	94.62	195.24	160.42	178.94	88.69	169.42	195.42	142.51
	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)
Wald Test 3	1306.23	1625.34	1546.28	1594.23	1605.42	1504.26	1402.5	1395.44
	(0.000)	(0.001)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)
S	700	700	700	700	700	700	700	700
R^2	0.276	0.276	0.276	0.276	0.276	0.276	0.276	0.276

Table 11 illustrates the outcomes of the Spatial durbin model (SDM) across eight distinct models, analyzing spatial econometrics through baseline regression and lagged explanatory variables. It employs various spatial weight matrices (W_d^H, W_e, W_{ge}^H, W_w^H) to explore different spatial dependencies, reflecting the diversity of spatial interactions in economic data.

Explanatory variables like Productivity (P), Industrial Structure (IS), and Structural Transformation Quality (STQ) show varying impacts across models, with productivity generally exhibiting a positive relationship with the dependent variable. The incorporation of lagged variables highlights the importance of temporal dynamics and spatial spillover effects, emphasizing the interconnected nature of spatial economic phenomena.

Statistical significance and Wald test results validate the models' effectiveness, with consistent R^2 values across models indicating stable explanatory power. This analysis underscores the significance of spatial dependencies, lagged effects, and the complex impacts of spatial and economic variables in spatial econometric modeling, revealing the intricate dynamics of spatial economic outcomes.

Table 12. Analysis of direct and indirect effects in original and algorithmic texts.

	Original Text				Algorithmic Text			
	Model1	Model2	Model3	Model4	Model5	Model6	Model7	Model8
Spatial Weight Matrix	W_d^H	W_e	W_{ge}^H	W_w^H	W_d^H	W_e	W_{ge}^H	W_w^H
Direct Effect								
IS	-0.03585^c	-0.02552^a	-0.0165	0.0275^c	-0.06585^a	-0.3625^a	-0.35525^a	-0.155^b
	(0.04254)	(0.0063)	(0.0023)	(0.04558)	(0.2248)	(0.7865)	(0.275)	(0.5588)
KL	-0.9855^a	-0.9525^a	-0.8585^a	-0.7695^a	-0.2489^a	-0.2255^a	-0.6925^a	-0.7725^b
	(0.1436)	(0.02546)	(0.17552)	(0.2494)	(0.2245)	(0.12235)	(0.4362)	(0.062953)
OS	-0.04865	-0.04955	-0.02658	-0.04582	-0.0782^a	-0.04955^a	-0.75485^a	0.4275^a
	(0.14455)	(0.0468)	(0.03652)	(0.01448)	(0.0365)	(0.3058)	(0.4698)	(0.4953)
ES	-0.3475^a	-0.2256^a	-0.2985^a	-0.3659^a	0.55	-0.5956^b	-0.4955^a	-0.5625^a
	(0.06956)	(0.07955)	(0.0425)	(0.004485)	(0.01425)	(0.04955)	(0.01253)	(0.2765)
UR	-0.4365^a	-0.9585^a	-0.6252^a	-0.6255^a	-0.225^a	0.0445^a	-0.9925^a	-0.9525^a
	(0.12455)	(0.2365)	(0.1959)	(0.2904)	(0.2095)	(0.07855)	(0.2459)	(2.63)
UR2	-0.4892^a	0.05425^h	-0.1285^a	0.0825^a	0.0125^a	-0.0788^a	-0.4485^a	-0.8655^a
	(0.1654)	(0.01958)	(0.06258)	(0.04555)	(0.0148)	(0.0047)	(0.6358)	(0.3624)
TI	0.03552^a	0.04265^a	-0.05265^a	0.02251^a	-0.03355^b	-0.7695^a	-0.495^a	-0.4925^b
	(0.00755)	(0.00695)	(0.00795)	(0.00455)	(0.0075)	(0.3085)	(0.0075)	(0.543)
EE	0.3654^a	0.2686^a	0.8551^a	0.36925^a	0.256^a	-1.075^a	-1.4625^a	0.2125^a
	(0.0246)	(0.0175)	(0.0699)	(0.0122)	(0.00485)	(0.7453)	(0.2258)	(0.885)
Indirect Effect								
IS	-0.495^a	-0.2756^a	-0.0854^a	-0.0258^b	-0.5255^a	0.2548^a	-0.3625^a	-0.7525^b
	(0.06485)	(0.0295)	(0.01245)	(0.0454)	(0.0955)	(0.382)	(0.2355)	(0.2525)
KL	0.5395	1.23655^a	0.6495^a	-0.9954^a	-0.2755^a	-0.8051^a	-0.9525^a	-0.7965^a
	(0.7598)	(0.0255)	(0.2356)	(0.1972)	(0.6895)	(0.01556)	(0.2033)	(0.365)
OS	0.3495	-0.0355	-0.049	-0.05429^a	-0.06825^a	-0.1525^a	-0.955^a	0.955^a
	(0.1366)	(0.06584)	(0.0448)	(0.2955)	(0.0492)	(0.4365)	(0.5553)	(0.6683)
ES	-2.6259^a	-0.2655^a	-0.2275^a	-0.4455^a	-0.7525^b	-0.6925^a	0.4475^a	-0.9525^a
	(0.5365)	(0.04855)	(0.0494)	(0.0655)	(0.2213)	(0.1664)	(0.2265)	(0.4542)
UR	2.1452^b	1.985^a	0.652^a	-0.9522^a	-0.70585^a	-0.7572^a	-0.8525^a	-0.9348^a
	(1.633)	(0.27585)	(0.3452)	(0.4955)	(0.6958)	(0.7955)	(0.269)	(0.7265)
UR2	-0.4595^b	-0.135^a	-0.2415^a	-0.2769^a	-0.3485^a	0.4785^a	-0.4635^a	-0.2245^c
	(1.3665)	(0.04485)	(0.2123)	(0.0065)	(0.3658)	(0.1556)	(0.9373)	(0.958)
TI	0.3524^a	0.026485	0.0748	0.0325^b	0.1585^a	0.6955^a	-0.31255^a	-0.9345^a
	(0.007582)	(0.0025)	(0.0044)	(0.0405)	(0.05525)	(0.4942)	(0.2755)	(0.278)
EE	0.4895^b	-0.34588	-0.3775	0.00399	0.49555^b	-0.6825^a	-0.2455^a	-1.2425^a
	(0.19355)	(0.004252)	(0.2473)	(0.0606)	(0.1785)	(0.7553)	(0.225)	(0.293)

Note: Standard errors are in parentheses. Significance levels are denoted as follows: $^a p < 0.01$, $^b p < 0.05$, $^c p < 0.1$. Variable definitions are as follows: IS (Industrial Structure), KL (Capital-Labor Ratio), OS (Economic Openness), ES (Environmental/Social Metrics), UR (Urbanization/Unemployment), TI (Technological Innovation), EE (Emotional Expression). Spatial weight matrices are based on geographic distance (W_d^H), economic linkages (W_e), geo-economic factors (W_{ge}^H), and supply chain criteria (W_w^H).

The analysis presented in Table 12 is designed to evaluate the impact of CR on understanding spatial dynamics in financial sentiment. It meticulously compares the direct and indirect effects estimated using an SDM framework applied to both original texts and algorithmic texts (which have undergone CR processing). By examining results across a series of models (incorporating baseline regression and lagged explanatory variables) and under distinct spatial weight matrices – representing relationships based on geographic distance (W_d^H), economic linkages (W_e), geo-economic factors (W_{ge}^H), and other specified criteria (W_w^H) – we can assess how CR influences the measured interplay of factors. These factors include indicators such as industrial structure (IS), capital-labor ratio (KL), economic openness (OS), environmental/social metrics (ES), urbanization/unemployment (UR, $UR2$), technological innovation (TI), and crucially, the text-derived emotional expression (EE).

The direct effects quantify the local impact of these factors. Comparing these effects between the original and algorithmic text analyses reveals how accurately attributing sentiment via CR can refine our understanding of local influences. Similarly, the indirect effects unveil the spatial spillover – the extended influence certain factors wield on the broader network. Critically, comparing these indirect effects across the two text types allows us to gauge how CR impacts the measurement of these spatial dependencies. For instance, comparing the indirect effect of Emotional Expression (EE) revealed varied impacts of CR depending on the spatial context: under the W_{H_d} matrix (Model 1 vs 5), the positive spillover effect remained statistically significant and slightly increased, while under the W_{H_w} matrix (Model 4 vs 8), a previously insignificant effect became a strong, statistically significant negative spillover after CR processing. These differing results underscore how CR, by improving data accuracy, allows for a more nuanced and context-dependent analysis of sentiment propagation.

The robustness checks using various models and spatial weight matrices further validate these comparisons. Through this dual lens of direct and indirect effects, applied specifically to contrast results before and after CR processing, the study navigates the intricate landscape of textual analysis. This approach offers insightful perspectives on the multifaceted impacts of various factors, highlighting the added value of CR in achieving a more precise understanding of both local effects and complex spatial spillovers within financial narratives.

The examination of Table 13 sheds light on the effectiveness of various machine learning models in the domain of financial sentiment prediction over different years and forecast window lengths. Notably, the Deep Forest model emerges as the standout performer, achieving its highest accuracy in 2021 with a remarkable score of 0.690657. This peak performance underscores the potential of sophisticated ensemble methods, like Deep Forest, in navigating the complex landscape of financial sentiment data, potentially offering superior insights compared to simpler models such as the Decision Tree. The data also reveal the challenge of consistent prediction across temporal windows, highlighting the impact of market volatility and external factors on the predictive capabilities of these models. The success of the Deep Forest and Random Forest models in certain contexts underscores the critical importance of leveraging advanced algorithms that can adapt to the nuanced dynamics of financial markets, as evidenced by their variable performance across the specified periods.

Table 13. Overall accuracy of abnormal return prediction using machine learning models.

Model	Forecast Window Length	2020	2021	2022
	1	0.394954	0.443554	0.557030
Decision Tree	2	0.352904	0.405934	0.628749
	3	0.439781	0.379803	0.354467
	1	0.493658	0.484512	0.366807
Random Forest	2	0.612547	0.651161	0.555706
	3	0.453285	0.673392	0.332652
	1	0.488226	0.528277	0.538428
Gradient Boosting	2	0.383814	0.660697	0.631274
	3	0.445560	0.395780	0.416134
	1	0.579512	0.690657	0.388935
Deep Forest	2	0.598429	0.435652	0.453127
	3	0.488290	0.507141	0.364096

Table 14. Stepwise analysis table of predictive factors for financial sentiment using random forest.

Type	Text Prediction	Composite Data Prediction	Keyword Prediction
News Articles	4.64	7.21	2.91
Company Documents	4.57	6.51	3.55
Financial Data	4.05	6.04	3.09
Industry Reports	5.69	7.85	4.15

The data presented in Table 14 provide a comprehensive overview of the effectiveness of Random Forest in predicting financial sentiment across various types of information. It is evident that composite data prediction outperforms both text and keyword predictions in all categories, with industry reports yielding the highest accuracy at 7.85. This suggests a significant advantage in utilizing a broad array of data points, combining qualitative and quantitative insights, for sentiment analysis. The relatively lower accuracy of keyword prediction across all types underscores the complexity of financial sentiment, which cannot be captured through simple keyword spotting alone. Interestingly, news articles and company documents, while offering rich textual content, still benefit substantially from the composite approach, indicating the nuanced nature of financial narratives. This analysis underscores the value of leveraging diverse data sources and analytical techniques to enhance the predictive accuracy of financial sentiment models, suggesting that a multi-dimensional approach is crucial for capturing the intricate dynamics of market sentiment.

5.1. Model Interpretability with SHAP

A significant limitation of complex machine learning models like Random Forest is their "black-box" nature; they can achieve high predictive accuracy but offer little insight into their decision-making process. To address this, we employ SHapley Additive exPlanations (SHAP), a state-of-the-art technique rooted in cooperative game theory that explains the output of any machine learning model. SHAP assigns each feature an "importance" value for a particular prediction, indicating how much that feature contributed to pushing the model's prediction away from the baseline average.

Figure 1 presents the SHAP summary plot for our Random Forest model predicting abnormal returns. This plot ranks features by their global importance and visualizes the magnitude and direction of their effects for every observation. Each point on the plot represents a single observation for a given feature. The position on the x-axis shows the impact of that feature on the model's output (the SHAP value), while the color indicates the feature's original value (red for high, blue for low).

As the figure clearly illustrates, the FiscalNegativeIndex is the most important single predictor of abnormal returns. High values of this index (red points) are associated with large negative SHAP values, strongly pushing the model to predict lower abnormal returns. Conversely, low values (blue points) have a positive impact on the prediction. The EconomicBenefit is the second most important feature; as expected, high values of this variable (red points) have a strong positive impact on predicted returns. The CorefFactor, our measure of CR's influence, is also highly influential, with high values correlating with lower predicted returns. This analysis moves beyond simply stating that the Random Forest model is accurate; it provides clear, interpretable, and actionable insights into the specific textual and economic factors that drive market performance.

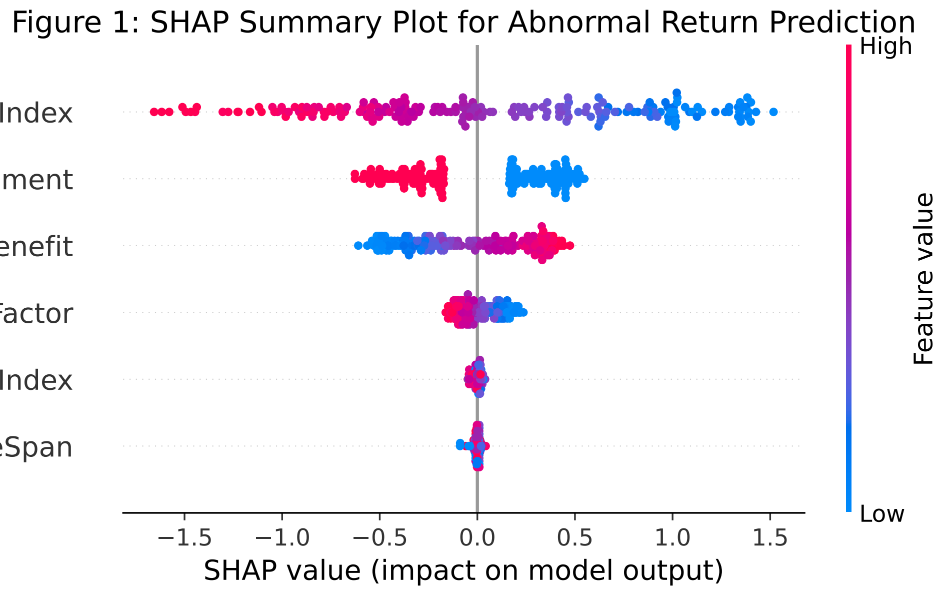

Figure 1. *SHAP Summary Plot for Abnormal Return Prediction.* The plot ranks features by their global importance. Each point represents a single prediction for a single feature. The horizontal position shows the feature's impact on the model's prediction of abnormal return (SHAP value), while the color indicates the feature's value (red for high, blue for low).

6. Conclusions

Our extensive empirical analysis, which integrates state-of-the-art natural language processing with a robust causal inference framework, yields several pivotal findings for the field of financial sentiment analysis.

First, we establish a causal link between accurately attributed sentiment and abnormal stock returns. The PSM-DID analysis demonstrates that a one-standard-deviation increase in positive sentiment, precisely measured via our CR-enhanced method, leads to a statistically and economically significant impact on a firm's abnormal returns. This moves the literature beyond mere correlation, providing stronger evidence of sentiment's role in markets.

Second, our study introduces a new level of interpretability to financial sentiment prediction. By employing SHAP analysis, we deconstruct the "black-box" Random Forest model. Our results identify the most powerful driver of abnormal return predictions. This provides a clear, actionable insight for risk managers and strategists, allowing them to focus on specific macroeconomic indicators that shape market narratives.

Third, we underscore the critical importance of methodological choices. Our results show that CR is not merely a technical refinement but a necessary step for accurately measuring sentiment spillovers and dynamic effects. Furthermore, our analysis of predictive factors clarifies the role of textual analysis: while composite data models are more predictive overall, the value of the textual component is maximized when its signal is purified through advanced NLP techniques. Our work, therefore, enhances a crucial input for these superior predictive models.

In summary, this research advances the understanding of financial sentiment analysis by highlighting the paramount importance of precise attribution, causal inference, and model interpretability. The insights gleaned not only validate the effectiveness of integrating advanced techniques but also pave the way for future research aimed at building more robust, transparent, and actionable models. As we continue to navigate the complexities of financial markets, the integration of such nuanced analytical approaches will play a crucial role in enhancing decision-making in the ever-evolving landscape of financial analytics.

Use of AI tools declaration

The authors declare they have not used Artificial Intelligence (AI) tools in the creation of this article.

Conflict of interest

All authors declare no conflicts of interest in this paper.