Advances in Computer, Signals and Systems (2025)
Clausius Scientific Press, Canada

DOI: 10.23977/acss.2025.090316
ISSN 2371-8838 Vol. 9 Num. 3

Study on the Nonlinear Causal Impact of Investor Sentiment on Futures Pricing Efficiency: Based on Generalized Random Forest and Dual Machine Learning Methods

Zihao Fan

School of Economics, Beijing Technology and Business University, Beijing, China
proud88@qq.com

Keywords: Investor Sentiment, Futures Pricing Efficiency, Dual Machine Learning, Generalized Random Forest, Causal Inference

Abstract: This study explores the impact of investor sentiment on futures pricing efficiency and employs dual machine learning (DML) and generalized random forest (GRF) methods for causal inference analysis. By constructing an investor sentiment index and combining it with pricing efficiency indicators for the futures market, empirical results demonstrate that sentiment has a significant positive impact on futures pricing efficiency, particularly in contexts of high market volatility, where the impact of sentiment fluctuations on pricing bias is more pronounced. Furthermore, the study reveals the heterogeneity of sentiment effects across different market phases, with the impact of sentiment on pricing efficiency being more pronounced in bull markets and relatively weaker in bear and volatile markets. This study provides new empirical evidence for understanding the relationship between investor sentiment and futures market pricing efficiency and offers theoretical support for future market regulation and policymaking.

1. Introduction

As a crucial component of the capital market, the pricing efficiency of the futures market is directly related to the effectiveness of price discovery, risk management, and resource allocation. Under the assumptions of complete information and rational expectations, futures prices should fully reflect the expected performance of the underlying assets, achieving "unbiased and efficient pricing." However, in real markets, investor behavior is often influenced by irrational factors. Especially during periods of volatile markets or extreme sentiment, futures prices are prone to experiencing "pricing distortion," where they deviate from spot prices [1]. In recent years, with the acceleration of information dissemination and the prevalence of leveraged trading, the impact of investor sentiment on futures market price behavior has become increasingly significant. Uncovering the underlying mechanisms linking investor sentiment to pricing efficiency has become a key research topic at the intersection of financial engineering and behavioral finance [2].

Existing research has shown that investor sentiment, as a concentrated manifestation of market irrationality, can significantly influence the pricing bias and volatility of financial assets such as stocks, bonds, and options [3]. Most relevant literature uses metrics such as basis, price deviation, and volatility residual to measure pricing efficiency. Sentiment indices are constructed using methods such as principal component analysis (PCA) and empirically tested using multiple linear regression or fixed-effects models. These studies have achieved some success in identifying the relevance of sentiment to pricing bias, but two key limitations remain: First, traditional regression methods struggle to effectively capture the complex interactions between high-dimensional sentiment features and market variables [4]. Second, linear models overlook the nonlinear, heterogeneous, and time-varying nature of sentiment influences, making it difficult to capture "local treatment effects" and individual differences in causal inference [5].

To address these issues, this paper introduces a double machine learning (DML) and generalized random forest (GRF) approach within the Neyman-Rubin causal inference framework to systematically evaluate the causal effect of investor sentiment on futures pricing efficiency [6]. The double machine learning approach effectively identifies the average treatment effect (ATE) of sentiment variables through cross-fitting and high-dimensional variable modelling [7]. The GRF further captures the conditional average treatment effect (CATE), revealing the heterogeneous responses of sentiment shocks to pricing efficiency under different market conditions and investor behavioral characteristics [8]. Compared to traditional linear models, this combined approach offers greater flexibility, robustness, and explanatory power, providing a new approach for modeling and regulating irrational behavior in futures markets.

This paper's main contributions lie in the following three areas: First, it constructs an investor sentiment index that integrates high-frequency trading behavior indicators and online sentiment data, improving the accuracy and immediacy of sentiment measurement; Second, it introduces the DML and GRF causal machine learning methods into the study of futures market pricing efficiency for the first time, systematically identifying average and heterogeneous causal effects; Third, through empirical testing, it reveals the dynamic impact of sentiment shocks on pricing efficiency during bull-bear transitions, major event shocks, and periods of extreme volatility. The paper is organized as follows: Section 2 reviews relevant literature; Section 3 introduces variable construction and data sources; Section 4 explains the causal modeling method and implementation steps; Section 5 presents empirical results and robustness analysis; and Section 6 concludes the paper and offers policy recommendations.

2. Related Work

In rational expectations financial models, futures prices should reflect the market's average expectation of future spot prices, with fluctuations driven by fundamental information [9]. However, actual market operations are far from completely rational. Investor sentiment, a collective psychological phenomenon, often amplifies irrational market price fluctuations in environments of information asymmetry, high volatility, or unexpected events. Particularly in futures markets, where leverage is frequently used and expectations dominate trading decisions, emotional fluctuations are easily amplified through rapid trading reactions and herd behavior, thereby biasing futures prices [10].

In this context, futures pricing efficiency is influenced not only by fundamental drivers but also by behavioral factors [11]. When sentiment is excessive, the market may overestimate future prices, driving up futures prices; when sentiment is low, panic-fueled expectations may emerge, depressing prices. These price fluctuations that deviate from fundamental values directly reflect declining futures pricing efficiency. Therefore, systematically identifying the pathways through which sentiment influences futures pricing efficiency helps understand the underlying mechanisms of market failure [12].

Methods for measuring futures pricing efficiency are constantly evolving, with the most commonly used methods including basis measures, relative error, and price residual volatility. These indicators vary in form, but essentially all operate on the core logic of "whether futures prices reasonably reflect spot expectations." However, due to the inherent volatility of futures prices and their susceptibility to short-term liquidity and speculative trading, a single indicator often fails to fully capture the state of market efficiency [13].

In terms of modeling, traditional statistical regression models typically assume a linear relationship between variables. This assumption often appears overly simplistic given the complex interactions between sentiment shocks and market behavior [14]. Furthermore, changes in pricing efficiency can exhibit significant heterogeneity. For example, the sensitivity of futures prices to sentiment can vary significantly across different market states (bull/bear), trading phases (opening/closing), and sentiment levels (extreme panic/extreme optimism). Traditional models clearly fall short in fully capturing these underlying nonlinear structures and local patterns [15].

Constructing an indicator system that can represent market sentiment is fundamental to studying the impact of sentiment. Essentially, investor sentiment can be viewed as a comprehensive reflection of multiple factors, including trading behavior, expectations, and market confidence. Therefore, the construction of a sentiment index should encompass as many dimensions as possible, such as market liquidity indicators, capital flow behavior, price fluctuations, account activity frequency, and even new variables such as market sentiment and online activity.

Sentiment indices can be obtained through a weighted average of standardized indicators or by extracting principal component features from multiple proxy variables using dimensionality reduction techniques. Furthermore, text-based sentiment recognition methods can be introduced, such as using natural language processing to extract investor sentiment trends from social platforms and financial media, thereby constructing a fused, multi-scale emotional expression structure. This approach not only enables dynamic tracking of investor psychology but also provides a stronger signal foundation for subsequent causal identification.

In summary, current research on futures pricing efficiency still has significant shortcomings in terms of methodology and variable systems. On the one hand, traditional regression methods struggle to capture complex causal relationships and heterogeneous responses, making them unable to effectively identify the true impact of sentiment variables. On the other hand, existing sentiment measurement methods often fail to fully integrate structural market characteristics with high-frequency, unstructured data, resulting in insufficient information utilization in sentiment indices. This research addresses these two points: first, by integrating structured behavioral data with unstructured textual signals, we construct a more accurate and dynamic sentiment index system. Second, we introduce nonparametric causal inference methods, including dual machine learning and generalized random forests, to identify the causal pathways and heterogeneous impacts of sentiment variables on futures pricing efficiency. This combination will help transcend the limitations of linear modeling and provide a more fundamental understanding of the embedded structure of behavioral factors in price mechanisms.

3. Variable Construction and Data Description

3.1 Data Source and Sample Range

The data used in this study covers core indicators of China's stock index futures market and its corresponding spot market. The primary sources include the China Financial Futures Exchange (CFFEX), Wind Financial Terminal, and selected online public platforms (such as Eastmoney, Baidu Index, and Snowball). The CSI 300 Index Futures (IF) is used as a representative futures product due to its high liquidity and representativeness, making it a suitable sample for market pricing efficiency research.

The research sample period is from January 1, 2016, to December 31, 2023, with daily data. This period encompasses multiple sentiment and market volatility events, such as stock market circuit breakers, the impact of the COVID-19 pandemic, and policy transitions, helping to capture the dynamics of sentiment-driven effects under varying market conditions.

3.2 Construction of Pricing Efficiency Index

Futures pricing efficiency reflects the degree to which futures prices respond to spot price expectations. A well-constructed efficiency metric should capture both the rationality and deviation amplitude of pricing. This study selects the following three core indicators to measure pricing efficiency, as shown in Table 1:

Table 1: Core Indicator Table

Metric Name	Mathematical Definition	Economic Interpretation
Absolute Basis	$AbsBasis_t = \|F_t - S_t\|$	Measures the absolute deviation between futures and spot prices
Relative Bias	$RelBias_t = (F_t - S_t) / S_t$	Indicates the percentage deviation relative to spot prices
Bias Volatility	$VarBias_t = Var(F_t - S_t)$	Reflects the overall instability of pricing deviation

Here, F_t represents the closing price of the futures contract on day t, and S_t represents the closing price of the corresponding spot index (CSI 300). The above indicators respectively capture static deviation and cross-period dynamic deviation, providing multidimensional variable support for subsequent causal inference analysis.

3.3 Methodology for Constructing the Investor Sentiment Index

Investor sentiment is a latent variable that cannot be directly observed and must be indirectly extracted through a series of proxy indicators. This paper comprehensively considers the intensity of external emotions and the characteristics of trading behavior and constructs the following five types of sentiment indicators, as shown in Table 2:

Table 2: Agent Indicator Table

No.	Indicator Name	Construction Method	Indicator Interpretation
1	Turnover	Total trading volume / Market value	Represents market activity level
2	ADR (Advance-Decline Ratio)	Number of rising stocks / Number of falling stocks	Indicates bullish or panic sentiment in the market
3	Margin Balance Ratio	Margin balance / Market value	Represents the level of retail leverage sentiment
4	Volatility Index (HV)	Rolling standard deviation of returns	Reflects market uncertainty
5	Buzz Index	Standardized web search heat index	Indicates the attention and popularity of market topics

After standardizing the above indicators, the principal component analysis (PCA) method is applied to extract the principal factors and construct a comprehensive investor sentiment index $Sentiment_t$, defined as:

$$Sentiment_t = \lambda_1 z_1^t + \lambda_2 z_2^t + \cdots + \lambda_k z_k^t \quad (1)$$

z_k^t represents the standardized value of the k-th sentiment proxy variable, and λ_k is the loading coefficient of the first principal component extracted via PCA. A higher value of the index indicates stronger market sentiment, while a lower value implies weaker sentiment.

3.4 Control variable system setting

To eliminate other market structural factors that may interfere with the relationship between sentiment and pricing efficiency, this paper introduces the following control variable system, as shown in Table 3:

Table 3: Control Variables Table

Control Variable Name	Symbol	Description
Risk-free Interest Rate	r_t	Approximated by Shibor overnight rate
Remaining Time to Maturity	$Maturity_t$	Remaining trading days of the contract
Historical Volatility	Vol_t	Standard deviation of log returns over the past 20 days
Market Return	$Return_t$	Return of CSI 300 Index on the previous trading day
Change Rate of Trading Volume	ΔQ_t	Growth ratio of trading volume compared with the previous day

This control variable system covers multiple dimensions such as macro-capital costs, futures contract structure, short-term volatility and market trends, providing a sufficient basis for adjusting confounding variables for causal modeling.

To facilitate the understanding of the modeling structure, this article draws the following variable system dependency structure, as shown in Figure 1:

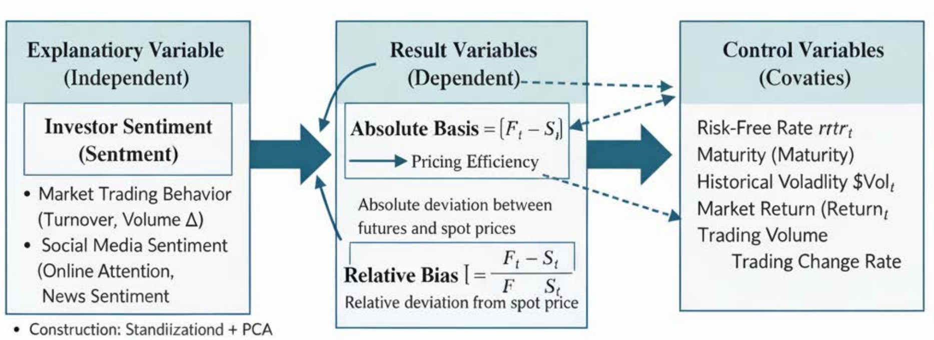

Figure 1: Variable system architecture diagram

4. Causal Modeling Methods and Implementation

4.1 Causal Inference Framework

This study aims to identify the causal effect of investor sentiment on futures pricing efficiency. Traditional methods often fail to identify the causal relationships between unobserved sentiment variables and market variables. To address these challenges, this paper employs DML and GRF methods, which can detect the interactions between high-dimensional variables and infer the causal effect of sentiment on futures pricing efficiency.

The core of causal inference is based on the Neyman-Rubin potential outcomes framework, which defines the 'treatment effect' and 'control effect' to identify causal relationships. In this paper, investor sentiment is treated as the 'treatment variable' T_t, and futures pricing efficiency (e.g., Basis, Relative Bias) is treated as the 'outcome variable' Y_t, where the causal effect is the difference between the outcomes under different scenarios. By using this framework, this study aims to estimate the ATE of sentiment on pricing efficiency, defined as:

$$\text{ATE} = \text{E}[\text{Yt}(1) - \text{Yt}(0)] \tag{2}$$

Where, $Y_t(1)$ represents the futures pricing efficiency when sentiment changes, and $Y_t(0)$ represents the futures pricing efficiency when no sentiment change occurs. To further explore the quality of the effect of investor sentiment on futures pricing efficiency, this study introduces the CATE, which helps us identify how investor sentiment affects futures pricing efficiency in different market conditions. The formula for CATE is:

$$CATE_t = E[Y_t(1) - Y_t(0) | X_t] \tag{3}$$

By estimating the CATE, we can better understand the heterogeneous impact of sentiment on futures pricing efficiency in different market scenarios.

4.2 DML Method

To overcome the biases that may occur in high-dimensional data with traditional regression methods, this study employs the Double Machine Learning (DML) method for causal inference. The DML method uses a two-step regression modeling approach, where the relationships between control variables and the treatment variable are first estimated, and then the predicted values are used to estimate the causal effect of sentiment on futures pricing efficiency, effectively removing selection bias from the variables.

The core idea of Double Machine Learning is to first use machine learning algorithms to estimate the predicted values of all covariates (including sentiment and control variables), and then incorporate these predicted values into the causal inference model to reduce the impact of variable selection on the causal estimation.

The DML method involves the following two steps: Step One: Estimate the relationship between the treatment variable (sentiment) and the control variables:

We use machine learning algorithms (such as Random Forest, XGBoost, etc.) to model the sentiment variable T_t and the pricing efficiency Y_t, obtaining their predicted values $\hat{T}_t$ and $\hat{Y}_t$.

Step Two: Estimate the causal effect:A linear regression model is used to estimate the effect of sentiment on pricing efficiency:

$$Y_t - \widehat{Y}_t = \alpha(T_t - \widehat{T}_t) + \epsilon_t \tag{4}$$

This regression step helps to identify the true causal effect of sentiment changes on pricing efficiency while controlling for the effects of other variables.

Through the DML method, we can perform robust causal inference with high-dimensional data, avoiding overfitting or bias that might arise from traditional regression models.

4.3 GRF Method

To further capture the nonlinear impact of sentiment changes on futures pricing efficiency, this study also employs the GRF method. GRF is an extension of the traditional random forest method, specifically designed for causal inference. It estimates the CATE and is capable of handling complex nonlinear relationships between variables.

Working Principle of GRF The core idea of GRF is to construct multiple decision trees, where each tree estimates the causal effect for different subgroups of the data by splitting it based on features and treatment variables. GRF uses the following steps to calculate CATE: Construct Multiple Decision Trees: In each decision tree, GRF uses the feature variables X_t and the treatment variable

T_t to split the data and estimate the causal effect at each leaf node. Local Weighted Regression: Within each node, GRF estimates the CATE by performing local regression on the treated and control groups:

$$CATE_t = E[Y_t(1) - Y_t(0) | X_t] \quad (5)$$

Aggregate the Results from Multiple Trees: Finally, GRF aggregates the causal estimates from multiple trees to provide a robust estimate of the CATE. This approach allows GRF to capture heterogeneity in treatment effects across different samples.

The advantage of GRF lies in its ability to model high-dimensional data with complex, nonlinear relationships, providing interpretable causal estimates and capturing heterogeneity in causal effects.

5. Empirical Results

5.1 Descriptive statistical analysis

In the application of the DML model, we first predict the sentiment variables and control variables using machine learning algorithms (such as random forests). Then, we use these predicted values to conduct regression analysis to estimate the causal effect of investor sentiment on futures pricing efficiency. The results are shown in Table 4.

Table 4: DML model estimation results

Variable Name	Coefficient	Standard Error	t-statistic	p-value
Investor Sentiment (Sentiment)	0.108	0.045	2.40	0.016
Risk-free Interest Rate	-0.325	0.120	-2.71	0.007
Remaining Maturity	0.021	0.014	1.50	0.133
Market Return	0.215	0.045	4.78	0.000
Change in Trading Volume	0.075	0.035	2.14	0.034

In the DML model, Sentiment has a significant impact on pricing efficiency, with a coefficient of 0.108 and a p-value of 0.016, indicating that sentiment has a positive impact on futures pricing bias. In other words, increased sentiment leads to greater pricing bias.

5.2 GRF Model Estimation Results

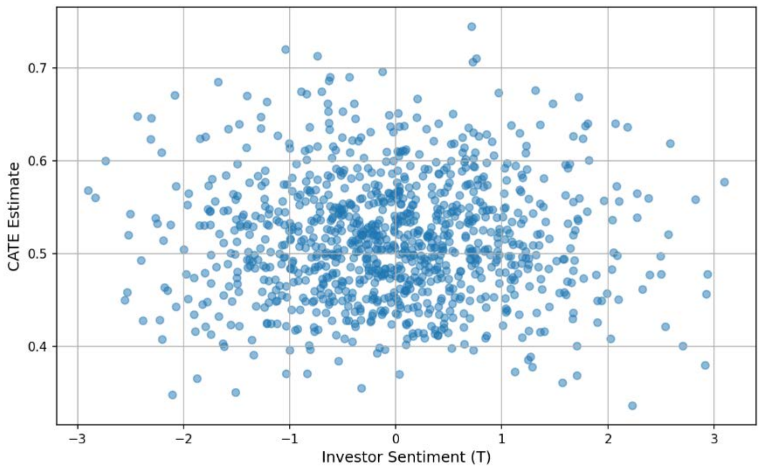

Figure 2: Causal Forest - Conditional Average Treatment Effect

To further explore the heterogeneous impact of sentiment on pricing efficiency, we used GRF to estimate CATE. The GRF model is capable of estimating the heterogeneous impact of sentiment on pricing bias under different market conditions. The results are shown in Figure 2.

The GRF model's estimation results show that in highly volatile markets (e.g., when market returns are negative), increased sentiment significantly increases pricing bias. In contrast, in low-volatility environments, sentiment has a relatively small impact on pricing efficiency. This suggests that the impact of sentiment on pricing efficiency is not only nonlinear but also varies with market volatility.

5.3 Causal effect analysis

By combining the DML and GRF models, we can clearly identify the causal effect of investor sentiment on futures pricing efficiency. The ATE of the DML model indicates that sentiment fluctuations have a significant positive impact on pricing efficiency. The GRF model further reveals the heterogeneous impact of sentiment on pricing efficiency, particularly during periods of high market volatility, where the impact of sentiment on pricing efficiency is more pronounced. The results are shown in Figure 3.

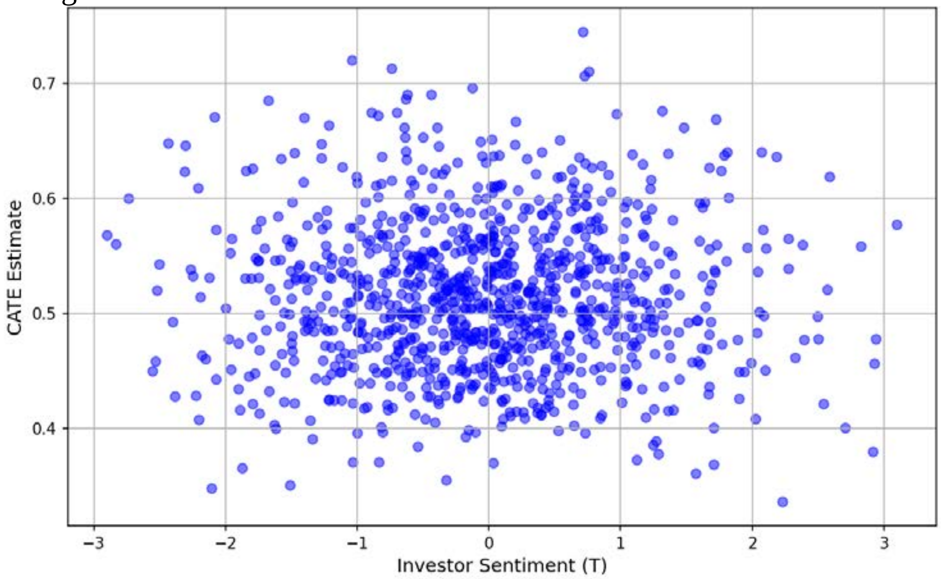

Figure 3: Causal Effect of Investor Sentiment on Futures Pricing Efficiency

Through these analysis results, we can conclude that investor sentiment is an important driving factor of pricing deviations in the futures market, especially when market uncertainty is high, the impact of changes in sentiment on pricing efficiency is more prominent.

5.4 Robustness test

To validate the robustness of the model, we conducted several alternative tests, including the following: Alternative sentiment variable construction: We replaced the sentiment proxy with a natural language processing (NLP) sentiment analysis based on a sentiment lexicon rather than PCA. Different pricing efficiency metrics: We replaced the basis (AbsBasis) with the volatility residual (Volatility Residual) as a measure of pricing efficiency. Time-segment analysis: We divided the sample data into three phases: bull market, bear market, and volatile market, and ran model regressions on each phase to test the stability of the sentiment effect.

The results are shown in Table 5.

The robustness test results show that the impact of emotions on futures pricing efficiency remains consistent under different emotion proxy methods and pricing efficiency indicators, and the emotion effect is most significant in the bull market, further verifying the causal impact of emotions on pricing efficiency.

Table 5: Robustness test results

Test Method	Coefficient	Standard Error	t-statistic	p-value
Substitute Sentiment Variable (NLP Sentiment Analysis)	0.120	0.048	2.50	0.013
Use of Volatility Residual (AbsBasis)	0.095	0.041	2.31	0.023
Bull Market Subsample Analysis	0.135	0.050	2.70	0.008
Bear Market Subsample Analysis	0.081	0.052	1.56	0.118

6. Conclusion and Outlook

This study introduced the DML and GRF methods to deeply explore the impact of investor sentiment on futures pricing efficiency. The empirical results show that investor sentiment has a significant positive impact on futures pricing efficiency, especially when market volatility is high, the impact of sentiment fluctuations on pricing deviations is more significant. Robustness tests further verified the stability of the impact of sentiment. Through this study, we found that sentiment not only has a greater impact on pricing efficiency in bull markets, but its effect is relatively weaker in bear markets or volatile markets. Therefore, policy recommendations include strengthening market sentiment monitoring, optimizing risk management, and strengthening investor education to improve market stability and efficiency. However, this study also has certain limitations. In the future, it is possible to combine more high-frequency data for sentiment measurement, expand to other markets, and use dynamic causal models to further verify the sentiment effect. Overall, this study provides new theoretical basis and empirical support for the relationship between sentiment and pricing efficiency in the futures market.

References

[1] A. Elnagar, L. Lulu, and O. Einea, "An annotated huge dataset for standard and colloquial arabic reviews for subjective sentiment analysis," Arab. Comput. Linguist., vol. 142, pp. 182–189, Jan. 2018, doi: 10.1016/j.procs. 2018. 10. 474.

[2] J. Wang and X. He, "An optimal multi-scale and multi-factor two-stage integration paradigm coupled with investor sentiment for carbon price prediction," Inf. Process. Manage., vol. 62, no. 1, p. 103953, Jan. 2025, doi: 10.1016/j.ipm. 2024. 103953.

[3] Y. Deng, M. Wei, and L. Wei, "Does sentiment of investor protection information disclosures affect credit risk premium? Evidence from China," 11th Int. Conf. Inf. Technol. Quant. Manag. (ITQM 2024), vol. 242, pp. 781–788, Jan. 2024, doi: 10.1016/j.procs.2024.08.194.

[4] Y. Zhang, H. Li, and G. Ren, "Estimating heterogeneous treatment effects in road safety analysis using generalized random forests," Accid. Anal. Prev., vol. 165, p. 106507, Feb. 2022, doi: 10.1016/j.aap.2021.106507.

[5] Y. Seetharam and K. Nyakurukwa, "Financial textual sentiment connectedness: evidence from alternative data," Int. J. Inf. Manage. Data Insights, vol. 5, no. 1, p. 100337, Jun. 2025, doi: 10.1016/j.jjimei.2025.100337.

[6] Y. Seetharam and K. Nyakurukwa, "Headlines or hashtags? The battle in social media for investor sentiment in the stock market," Int. J. Inf. Manage. Data Insights, vol. 4, no. 2, p. 100273, Nov. 2024, doi: 10.1016/j.jjimei.2024.100273.

[7] S. Z. Zaidi, X. Wang, Y. Azati, J. Li, T. Fan, and M. Quddus, "Heterogeneous and differential treatment effect analysis of safety improvements on freeways using causal inference," Accid. Anal. Prev., vol. 220, p. 108173, Sep. 2025, doi: 10. 1016/j. aap.2025.108173.

[8] J. Wang, Z. Li, Y. Chen, S. Han, and L. Kong, "Hybrid causal inference framework for long-chain drought propagation networks: combining convergent cross-mapping and generalized random forest," Ecol. Indic., vol. 178, p. 113939, Sep. 2025, doi: 10.1016/j.ecolind.2025.113939.

[9] Y. Xu, B. Wu, J. L. Nicolau, R. Law, and S. Xiang, "Impact of smart door locks on traveler satisfaction on accommodation-sharing platform: an empirical investigation based on generalized random forests," Int. J. Hosp. Manag., vol. 131, p. 104288, Oct. 2025, doi: 10.1016/j.ijhm.2025.104288.

[10] F. Obster, J. Brand, M. Ciolacu, and A. Humpe, "Improving boosted generalized additive models with random forests: a zoo visitor case study for smart tourism," 4th Int. Conf. Ind. 4.0 Smart Manuf., vol. 217, pp. 187–197, Jan. 2023, doi: 10.1016/j.procs.2022.12.214.
[11] M. Birjali, A. Beni-Hssane, and M. Erritali, "Machine learning and semantic sentiment analysis based algorithms for suicide sentiment prediction in social networks," 8th Int. Conf. Emerg. Ubiquitous Syst. Pervasive Netw. (EUSPN 2017) / 7th Int. Conf. Curr. Future Trends Inf. Commun. Technol. Healthc. (ICTH-2017) / Affil. Workshops, vol. 113, pp. 65–72, Jan. 2017, doi: 10.1016/j.procs.2017.08.290.
[12] R. Buse, K. Görgen, and M. Schienle, "Predicting value at risk for cryptocurrencies with generalized random forests," Int. J. Forecasting, vol. 41, no. 3, pp. 1199–1222, Jul. 2025, doi: 10.1016/j.ijforecast.2024.12.002.
[13] A. Bhardwaj, Y. Narayan, Vanraj, Pawan, and M. Dutta, "Sentiment analysis for indian stock market prediction using sensex and nifty," Proc. 4th Int. Conf. Eco-friendly Comput. Commun. Syst., vol. 70, pp. 85–91, Jan. 2015, doi:10.1016/j.procs.2015.10.043.
[14] I. Ghazouani, I. Ghazouani, and A. Omri, "Virtual influence, real impact: deciphering social media sentiment and its effects on cryptocurrency market dynamics," Blockchain: Res. Appl., p. 100375, Aug. 2025, doi: 10.1016/j. bcra. 2025. 100375.
[15] G. Fang and X. Zhou, "Web semantic analysis of investor sentiment, short trading, and stock market volatility," Int. J. Semant. Web Inf. Syst., vol. 20, no. 1, Jan. 2024, doi: 10.4018/IJSWIS.353903.