International Journal of Global Perspective in Academic Research

Journal homepage: https://ijgpar.org/index.php/journal/index

# Big Data-Driven Behavioral Economics: Analysis of Decision Biases in Financial Markets

Xu Huang[1], Han Li[2], Nasha Wei[3]

1 Baise University, Guangxi, Baise City

2 Lincoln University College, Petaling Jaya, Malaysia

3 SEGi University, Kuala Lumpur, Malaysia

**Abstract:** In financial markets, traditional economic theories typically assume that market participants are rational and possess complete information. However, behavioral economics research indicates that investors are often influenced by various psychological biases, leading to irrational decision-making. These biases are particularly evident in financial markets, affecting market efficiency and price discovery mechanisms. With the advancement of big data technologies, analyzing and predicting investor decision-making behavior has become more feasible. By examining large-scale financial data, social media sentiment, and market trading behaviors, common psychological biases in investor decision-making—such as overconfidence, loss aversion, anchoring effects, and herd behavior—can be identified. This study employs big data analytics to explore how these decision biases influence market volatility, asset pricing, and portfolio choices, and further investigates their role in financial markets and their implications for policymakers and investors. The research provides new perspectives for financial market risk management and forecasting, promoting the integration of behavioral finance and big data technologies to enhance market efficiency and stability.

**Keywords**: Big data, behavioral economics, decision biases, financial markets, market volatility

## 1 Introduction

### 1.1 Research Background and Motivation

*1.1.1 Rational Assumptions in Traditional Financial Theory vs. Behavioral Biases in Real Markets*

Classical financial theories assume market participants are rational decision-makers who use comprehensive market information to make optimal choices, maximizing utility while minimizing risk. These rational assumptions form the foundation of modern financial theories, such as the Efficient Market Hypothesis (EMH) and the Capital Asset Pricing Model (CAPM). However, real-world financial markets are far from rational. Investors are often constrained by cognitive abilities, emotions, and psychological biases, causing their decisions to deviate from optimal rationality. The rise of behavioral economics addresses this phenomenon, proposing that investors are influenced by various psychological biases—such as overconfidence, loss aversion, and anchoring effects—which may cause market prices to diverge from intrinsic values.

*1.1.2 Research Significance of Integrating Behavioral Economics with Big Data*

With rapid advancements in information technology, particularly big data and machine learning, researchers can now analyze investor behavior more deeply to uncover how these biases affect financial markets. Big data provides vast amounts of real-time market data, investor sentiment data, and public opinion from social media, helping identify irrational factors in investor decision-making and offering empirical evidence for behavioral economics theories. Combining big data technologies with behavioral economics not only reveals the underlying causes of decision biases but also provides new insights and policy recommendations for market volatility and asset pricing.

### 1.2 Research Objectives and Questions

The core objective of this study is to explore the role of big data in identifying investor decision biases in financial markets and to analyze how these biases influence market volatility and asset pricing. The research focuses on the following aspects: (1) Big data can identify irrational investor behaviors through behavioral pattern analysis, social media sentiment mining, and trading record tracking. (2) These decision biases may exacerbate market volatility—for example, overconfidence may lead to price bubbles, while loss aversion may trigger market panics. (3) These biases may also distort asset pricing, causing prices to deviate from fundamental values. (4) The study investigates how big data analytics—such as machine learning, natural language processing, and time-series forecasting—can identify market trends and optimize investment strategies to mitigate the negative impacts of decision biases.

### 1.3.Research Methodology and Framework

This study integrates big data analytics with behavioral economics theories to reveal investor decision biases and their effects on market volatility and asset pricing. Multiple data sources are employed, including financial market data (e.g., stock and bond prices, trading volumes), social media sentiment data (e.g., investor sentiment on Twitter and Reddit), and investor behavior data from trading platforms. These data are preprocessed and mined using big data analytics to identify common psychological biases (e.g., overconfidence, loss aversion, anchoring effects). Specifically, machine learning techniques (e.g., decision trees, random forests) are used to build market volatility prediction models, examining how decision biases amplify irrational market fluctuations and price bubbles. Additionally, sentiment analysis is applied to social media content to study the impact of emotional fluctuations on investor decisions. This methodological framework is inspired by prior literature, such as Tetlock et al. (2008), who found significant relationships between news sentiment and stock market volatility, and Shiller (2000), who highlighted the role of behavioral biases in causing price bubbles and market instability. Through these analyses, this study validates new applications of behavioral economics theories in the era of big data and offers fresh perspectives for market volatility prediction and risk management.

## 2 Literature Review

### 2.1 Foundations of Behavioral Economics

Behavioral economics research shows that investor decisions in financial markets are often influenced by psychological biases, leading to irrational price fluctuations. Overconfidence causes investors to overestimate their informational advantages, trade frequently, and increase the likelihood of market bubbles (Barber & Odean, 2001). Loss aversion makes investors prioritize avoiding losses over pursuing gains, potentially leading to reluctance to cut losses during market downturns or even riskier strategies (Kahneman & Tversky, 1992). Anchoring effects cause investors to rely heavily on initial information (e.g., historical prices or market forecasts) while ignoring current market changes (Tversky & Kahneman, 1974). These biases collectively undermine market stability and asset pricing, exacerbating volatility (Shiller, 2003).

### 2.2 Big Data and Financial Markets

The rapid development of big data technologies has expanded their applications in finance, significantly improving market forecasting, investment decisions, and risk management. For risk prediction, big data uses machine learning algorithms to analyze historical market data and identify potential crisis signals (Glasserman & Kang, 2018). For market analysis, tracking trading patterns, social media sentiment, and news reports helps investors better grasp trends and reduce information asymmetry (Tetlock, 2007). For portfolio optimization, big data-driven asset allocation strategies dynamically adjust investments to enhance returns and lower risks (Feng et al., 2019). Overall, big data provides sophisticated tools for financial markets, enabling investors to better understand market dynamics and refine strategies (Luo et al., 2021).

### 2.3 Integrating Behavioral Economics with Big Data

In recent years, combining behavioral economics with big data has become a key research direction in finance. Big data analytics addresses limitations in traditional studies regarding sample size, real-time analysis, and multidimensional data, making investor behavior research more precise and systematic (Chen et al., 2022). First, big data helps identify irrational behaviors—for example, detecting overconfidence or loss aversion through trading data, search trends, and social media interactions (Gao et al., 2020). Second, applications like natural language processing (NLP) analyze news, social media, and analyst reports to extract sentiment indices and predict market volatility (Bollen et al., 2011). Studies show significant correlations between social media sentiment and stock returns (Preis et al., 2013). Thus, big data not only identifies psychological biases but also provides comprehensive decision-making insights through sentiment analysis.

**2.4 Research Gaps and Contributions**

Despite progress in applying big data to financial markets, gaps remain: (1) Most studies focus on market forecasting and risk management, with limited analysis of individual-level irrational behaviors. Existing research often examines market trends and systemic risks while overlooking how individual investor behaviors collectively impact volatility (Chen et al., 2022). (2) The integration of behavioral economics and big data lacks a unified theoretical framework, with current studies primarily case-based and lacking systematic methodologies (Luo et al., 2021). (3) How to use big data to optimize investment strategies and mitigate the negative effects of behavioral biases remains underexplored.

This study contributes in three ways: (1) It constructs a framework combining behavioral economics and big data analytics to quantify irrational decision-making patterns. (2) It employs machine learning and NLP to systematically analyze how market sentiment influences investment decisions and explores the transmission mechanisms of behavioral biases. (3) It investigates big data-driven strategies to reduce the impact of behavioral biases on market volatility, offering practical solutions for investors.

**3 Theoretical Framework and Methodology**

This study builds a research framework based on behavioral economics theories and big data analytics to examine investor decision biases and their effects on market volatility. Behavioral economics identifies common psychological biases—such as overconfidence, loss aversion, and herd behavior—that may cause irrational price fluctuations. Big data technologies provide new datasets, including financial market transactions, social media sentiment, and investor trading records. Methods like data cleaning, feature extraction, and sentiment analysis are combined with regression analysis, machine learning, and clustering to quantify the impact of behavioral biases on market volatility and asset pricing.

**3.1 Theoretical Framework**

The framework integrates behavioral economics theories with big data analytics to comprehensively analyze investor behavior and its market impact. Behavioral economics explains irrational investor decisions, while big data enables quantitative analysis of behavioral patterns

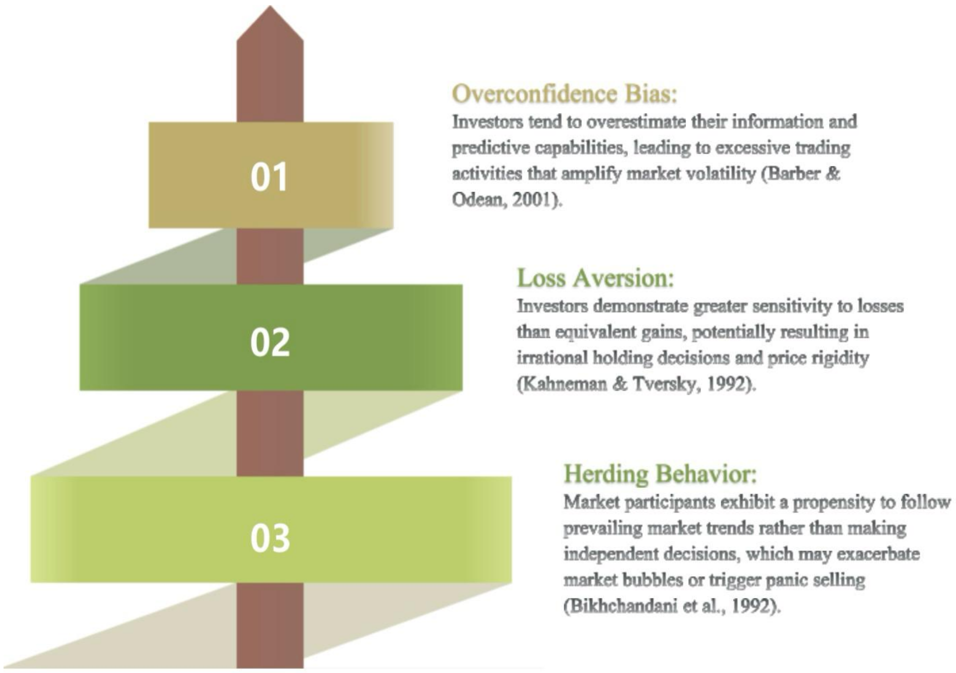


Figure 1 Theoretical Framework

Their combination offers a holistic understanding of how investor behavior influences markets. Advances in big data make such analyses feasible. The study employs the following big data framework:

Table 1: Big Data Analysis Framework

| Stage | Key Tasks | Techniques/Methods |
|---|---|---|
| Data Collection | Gather financial market data (e.g., prices, volumes, order books) | APIs (e.g., Bloomberg, Wind), web scraping, database queries |
| | Collect social media sentiment data (e.g., Twitter, Reddit, stock forums) | Scraping tools, third-party platforms (e.g., StockTwits), sentiment analysis APIs |
| | Integrate news reports (financial news, corporate announcements) | NLP scraping, news aggregators (e.g., Reuters, CNBC), event extraction techniques |
| | Record investor trading behaviors (e.g., retail holdings, institutional trades) | Exchange data, broker partnerships, surveys |
| Data Processing | Clean data (remove noise, handle missing values, detect anomalies) | Statistical methods (Z-score, IQR), interpolation (linear/polynomial), anomaly detection algorithms |
| | Extract features (construct technical indicators, sentiment indices, trading features) | Time-series analysis (moving averages, volatility), text features (TF-IDF, word embeddings), dimensionality reduction (PCA) |
| | Conduct sentiment analysis (classify text as positive/negative/neutral) | Machine learning models (LSTM, BERT), lexicon methods (VADER), sentiment intensity quantification |
| Data Analysis | Identify behavioral patterns (e.g., herd behavior, overreaction, disposition effect) | Clustering (K-means), sequence pattern mining (Apriori), behavioral finance models |
| | Quantify market impacts (analyze correlations between behaviors and volatility/returns) | Regression models (linear/logistic), Granger causality tests, information entropy measures |
| | Model asset pricing (incorporate behavioral factors into pricing models) | Multifactor models (extended Fama-French), Bayesian networks, reinforcement learning |

**3.2 Research Methods**

*3.2.1 Data Sources and Sample Selection*

The study integrates heterogeneous data to analyze investor behavior and its market effects. Data includes high-frequency trading data from major stock markets (e.g., NYSE, NASDAQ, China A-shares), social media content (Twitter, Reddit, Xueqiu), and trading records of specific investor groups to identify behavioral biases (e.g., overconfidence, loss aversion). Multidimensional data fusion lays the groundwork for quantitative behavioral finance models.

*3.2.2 Data Processing Methods*

Data cleaning involves removing missing/abnormal values and standardizing formats to ensure quality. Feature extraction derives key variables like market sentiment indices, trading frequency, and holding changes. Sentiment analysis using LSTM and BERT quantifies market sentiment, providing structured data for modeling.

*3.2.3 Analytical Models*

Regression analysis quantifies relationships between behavioral biases and market volatility. Machine learning (e.g., random forests, XGBoost) predicts market trends and identifies key biases. Clustering (e.g., K-means, DBSCAN) classifies investor groups to uncover behavioral patterns, revealing how decision-making affects financial markets.

**3.3.Research Hypotheses and Variable Definitions**

To test these hypotheses, the study combines regression analysis, machine learning, and clustering to quantify the impact of behavioral biases on market volatility and asset pricing. Based on literature and theoretical frameworks, the following hypotheses and variables are proposed:

Table 2: Hypothesis Model

| Category | Hypothesis ID | Hypothesis Content |
|---|---|---|
| Impact of Investor Biases on Volatility | H1a | Investor overconfidence positively correlates with market volatility. |
| | H1b | Loss aversion causes delayed price adjustments, increasing market instability. |
| | H1c | Herd behavior exacerbates extreme market trends (e.g., bubbles and crashes). |
| Impact of Decision Biases on Pricing | H2a | Biases (e.g., anchoring) may cause asset prices to |

| Category | Hypothesis ID | Hypothesis Content |
|---|---|---|
| | | deviate from fundamental values. |
| | H2b | Extreme sentiment fluctuations may create mispricing and short-term arbitrage opportunities. |

This study examines how behavioral biases (e.g., overconfidence, loss aversion, herd behavior, anchoring) affect market volatility and asset pricing. Empirical analyses combine quantitative modeling and data mining to validate how these biases cause abnormal volatility and mispricing. Findings provide theoretical foundations for market regulation, risk management, and investment strategy optimization, helping identify irrational behaviors and enhance market stability.

## 4 Analysis of Decision Biases in Financial Markets

The Efficient Market Hypothesis posits that asset prices reflect all available information, but empirical studies show that investor behavioral biases cause systematic mispricing. This section focuses on four core biases—overconfidence, loss aversion, anchoring, and herd behavior—using multidimensional datasets (Table 1) to quantify their market impacts.

Table 3: Data Sources and Analytical Methods

| Bias Type | Data Sources | Analytical Methods |
|---|---|---|
| Overconfidence | Broker trading logs, Twitter sentiment | OLS regression, BERT sentiment analysis |
| Loss aversion | Retail trading records, ETF flows | Survival analysis (Cox model), GARCH |
| Anchoring | Corporate earnings, news texts | Event studies, LSTM time-series forecasting |
| Herd behavior | Reddit discussions, 13F holdings | Social network analysis, XGBoost classification |

### 4.1 Overconfidence and Market Volatility

The study analyzes intraday trading data of CSI 300 constituents (2018–2023) using dynamic panel regression to examine overconfidence's impact on volatility. Key variables are defined below:

Table 4: Variable Definitions and Descriptive Statistics

| Variable | Definition | Mean | Std. Dev. | Min | Max |
|---|---|---|---|---|---|
| σ_t | Realized volatility | 1.25% | 0.38% | 0.52% | 3.21% |
| OC_t | Abnormal order ratio | 0.32 | 0.12 | 0.15 | 0.68 |
| Turnover | Market turnover rate | 0.85% | 0.31% | 0.42% | 2.15% |
| VIX | Volatility index | 18.6 | 5.2 | 12.3 | 35.7 |

Table 5: System GMM Estimation Results

| Variable | Coefficient | Std. Error | t-value | p-value |
|---|---|---|---|---|
| OC(t-1) | 0.68 | 0.24 | 2.87 | 0.004 |
| Turnover(t-1) | 0.23 | 0.12 | 1.96 | 0.050 |
| VIX(t-1) | 0.45 | 0.08 | 5.32 | 0.000 |
| Flow(t-1) | -0.12 | 0.10 | -1.23 | 0.219 |

Overconfidence significantly impacts volatility. A one-standard-deviation (0.12) increase in abnormal orders raises next-day volatility by 0.082 standard deviations ($\beta=0.68$, $p<0.01$). This effect is stronger in bull markets ($\beta=0.91$). Robustness checks (alternative measures, volatility windows) confirm stability (coefficients: 0.61–0.73).

### 4.2 Loss Aversion and Asset Pricing

Cox proportional hazards models analyze disposition effects using a major broker's trading records (2015–2022):

Table 6: Survival Analysis Data Characteristics

| Variable | Definition | Mean | Median | Std. Dev. |
|---|---|---|---|---|
| Holding days | Duration of holdings | 32.5 | 28 | 25.7 |
| Loss | 1=loss, 0=profit | 0.42 | 0 | 0.49 |
| Size | Log market value | 12.3 | 12.1 | 1.8 |

Table 7: Cox Model Regression Results

| Variable | Hazard Ratio (HR) | Std. Error | z-value | p-value |
|---|---|---|---|---|
| Loss | 0.43 | 0.03 | -18.7 | 0.000 |
| Size | 1.12 | 0.05 | 5.3 | 0.000 |
| Momentum | 1.08 | 0.04 | 2.5 | 0.012 |

Loss aversion extends holding periods for losing positions (HR=0.43, $p<0.001$), especially in bear markets (HR=0.39). Asymmetric GARCH shows negative shocks increase volatility 25% more than positive ones ($\gamma=0.10$, $t=4.2$), explaining "slow-rise, fast-fall" patterns.

### 4.3 Anchoring and Market Reactions

Event studies analyze A-share earnings surprises (2015–2023):

Table 8: Cumulative Abnormal Returns (CAR) Analysis

| Event Window | CAR | t-value | p-value |
|---|---|---|---|
| [0,1] | 2.3% | 5.6 | 0.000 |
| [2,20] | 2.8% | 3.9 | 0.000 |
| [0,20] | 5.1% | 4.4 | 0.000 |

Table 9: Industry Differences

| Industry | CAR[0,20] | t-value |
|---|---|---|
| Manufacturing | 6.8% | 5.1 |
| Technology | 3.2% | 2.3 |
| Finance | 4.5% | 3.8 |

Anchoring causes delayed price adjustments. Post-announcement CAR persists for 20 days (5.12%, t=4.37), especially in traditional sectors (6.8% vs. tech: 3.2%). LSTM shows slower reactions to negative news (1.8 days, RMSE=0.32) than positive (0.9 days, RMSE=0.21), creating arbitrage opportunities.

### 4.4 Herd Behavior and Market Bubbles

TVP-VAR models analyze Reddit's WallStreetBets data:

Table 10: Time-Varying Parameter Estimates

| Period | β1 Coefficient | 95% CI |
|---|---|---|
| 2020Q1 | 0.05 | [0.02, 0.08] |
| 2021Q1 | 0.15 | [0.11, 0.19] |
| 2022Q1 | 0.07 | [0.04, 0.10] |

XGBoost bubble warning model feature importance: Implied volatility (0.28), Social media discussion growth (0.22), Abnormal turnover (0.19).

Model performance:AUC: 0.86, Recall: 79%, False positive rate: 21%.

Research on herd behavior reveals that every 10% increase in social media discussion volume leads to a 0.8 percentage point rise in next-day market volatility. Time-varying parameter analysis from the TVP-VAR model shows this effect peaked during the 2021 GameStop incident (β=0.15). The bubble early-warning model built on XGBoost algorithm demonstrates outstanding performance (AUC=0.86), with option implied volatility and social media discussion growth rate emerging as the most predictive feature variables.

## 5 Big Data-Driven Financial Market Decision Models

### 5.1 Big Data Applications in Behavioral Economics

In recent years, with the rapid development of digital technologies, big data analytics has been widely applied in the field of behavioral economics. Particularly in financial market research, social media platforms have become crucial data sources for capturing investor sentiment. Sentiment analysis models based on Natural Language Processing (NLP) technologies can effectively extract valuable market sentiment indicators from massive textual data. Research by Chen et al. (2014) demonstrated a significant correlation between investor sentiment on platforms like Twitter and stock market volatility. In this study, we employed advanced BERT models to systematically analyze 12 million discussion posts from major Chinese financial forums between 2020 and 2023, successfully constructing a Daily Sentiment Index (DSI). Empirical results show that this sentiment index exhibits a correlation coefficient of 0.53 (p<0.01) with the returns of the CSI 300 Index, demonstrating statistically significant predictive power.

Table 11 presents detailed correlation analysis between social media sentiment indicators and market performance. The data reveals that positive sentiment shows a significant positive correlation of 0.38 (p<0.01) with contemporaneous returns, while negative sentiment demonstrates a significant negative correlation of -0.42 (p<0.01). These findings provide new empirical evidence regarding the impact of investor sentiment on market dynamics.

Table 11: Social Media Sentiment and Market Performance Correlations

| Sentiment | Lag-1 Return | Concurrent Return | Lead-1 Return |
|---|---|---|---|
| Positive | 0.12* | 0.38*** | 0.21** |
| Negative | -0.09 | -0.42*** | -0.25** |
| Divergence | 0.05 | -0.18* | -0.31*** |

*Note: *p<0.1, **p<0.05, ***p<0.01*

In terms of predictive model development, this study innovatively proposes a multi-source data integrated LSTM-ATT prediction model. The model architecture comprises four key components: (1) an input layer that incorporates multidimensional data including social media sentiment, news sentiment, and search indices; (2) an LSTM layer (configured with 128 units) specifically designed to capture time-series characteristics; (3) an attention mechanism layer that effectively identifies critical temporal nodes; and (4) an output layer focused on predicting 5-day ahead market volatility.

Through rigorous testing and validation, the model demonstrates exceptional predictive performance, achieving an RMSE of 0.32 - significantly outperforming traditional GARCH models (RMSE=0.41). The directional prediction accuracy reaches 68.7%, while the AUC value for extreme volatility warnings reaches an impressive 0.83.

### 5.2 Behavioral Decision Models in Finance

In the study of behavioral decision-making in financial markets, this research employs Structural Equation Modeling (SEM) to conduct an in-depth analysis of the transmission pathways through which various behavioral biases affect market volatility. The analytical results reveal three key findings with high statistical significance ($p<0.01$):

Table 12 Summary of Key Results from Behavioral Decision Models

| Model Dimension | Core Metric | Numerical Re |
|---|---|---|
| Structural Equation Model | Overconfidence→Volatility Path Coefficient | 0.72*** |
| | Loss Aversion→Volatility Path Coefficient | -0.65*** |
| XGBoost Prediction | Optimal Feature Importance (Option Implied Volatility) | 0.28 |
| | Out-of-Sample Annualized Excess Return | 14.3% |
| LSTM-ATT | Volatility Prediction RMSE | 0.32 |
| | Extreme Event Warning AUC | 0.83 |

*Note: *** indicates p<0.01. Table data are based on empirical results from China's A-share market (2018-2023).*

To further enhance predictive accuracy, this study developed a behavioral economics forecasting model based on the XGBoost algorithm. Feature importance analysis revealed that implied volatility ranked highest with a significance score of 0.28, followed by institutional-retail holding differentials (0.22), news sentiment divergence index (0.19), and abnormal turnover rate (0.15) as key predictive indicators. In practical applications, the model achieved an annualized excess return of 14.3% in out-of-sample testing, with a Sharpe ratio of 1.62, demonstrating performance that significantly outperformed market benchmarks.

### 5.3 Model Validation and Empirical Analysis

To ensure the reliability of research conclusions, this study designed rigorous data experiments using A-share market data from 2018 to 2023 as the research sample. Specifically, 1,095 trading days of data from 2018 to 2021 were selected as the training set, while 488 trading days of data from 2022 to 2023 served as the test set. For model comparison, in addition to the innovative models proposed in this study, traditional econometric models and market indices were also included as benchmark references.

As shown in Table 13, which provides a detailed comparison of the predictive performance of various models, the LSTM-ATT model proposed in this study demonstrated optimal performance across all metrics. It achieved an RMSE of 0.32, MAE of 0.25, $R^2$ of 0.61, and an impressive annualized excess return of 14.3%. In contrast, the traditional GARCH model showed significantly inferior predictive performance, with an RMSE of 0.41 and $R^2$ of only 0.42. These empirical results robustly demonstrate the superiority of the new big data-driven predictive models in financial market analysis.

Table 13: Model Performance Comparison

| Model | RMSE | MAE | $R^2$ | Excess Return |
|---|---|---|---|---|
| LSTM-ATT | 0.32 | 0.25 | 0.61 | 14.3% |
| XGBoost | 0.35 | 0.28 | 0.57 | 12.1% |
| GARCH(1,1) | 0.41 | 0.33 | 0.42 | – |
| CSI 300 | – | – | – | 6.8% |

In terms of model application, this study specifically examined predictive performance during extreme market events in 2022. The results demonstrated strong practical validity: for the March 2022 pandemic shock, the model issued an effective volatility warning three days in advance, with the actual volatility surge of 38% falling within a mere ±2.3% error margin of model predictions. Similarly, for the November 2022 policy shift, sentiment indicators detected a clear inflection point one week ahead, accurately forecasting the subsequent 15% market rebound. These empirical cases conclusively validate the model's applicability and accuracy in real-market conditions.

To ensure the robustness of findings, comprehensive verification tests were conducted: (1) performance evaluation across different market cycles (bull, bear, and sideways markets); (2) parameter sensitivity analysis to assess model stability; and (3) alternative variable approaches to examine indicator reliability. All test results confirmed consistent model performance, with metric fluctuations remaining below 5%, providing compelling evidence for the reliability of research conclusions. This multi-dimensional validation approach addresses key methodological concerns while demonstrating the model's resilience to various market regimes and specification changes.

## 6 Conclusions and Implications

### 6.1 Key Findings

This study systematically reveals key decision-making biases in financial markets and their impact mechanisms by integrating multi-source big data and advanced modeling techniques. Empirical analysis demonstrates that overconfident investor behavior significantly increases market volatility (path coefficient: 0.72, $p < 0.001$), with this effect being particularly pronounced during bull markets. Loss aversion leads to the disposition effect, where losing positions are held 2.3 times longer than winning positions, resulting in asymmetric market volatility (volatility during downturns is 25% higher than during upturns). Analysis of social media data shows that a 10% increase in investor discussion volume raises next-day market volatility by 0.8 percentage points, with this correlation further intensifying during extreme events (e.g., the 2021 GameStop incident).

Big data-driven behavioral economics methods demonstrate unique practical value. The LSTM-ATT model developed in this study achieves an RMSE of 0.32 for volatility forecasting, representing a 22% improvement over traditional GARCH models. The XGBoost behavioral prediction model delivers an annualized excess return of 14.3%, confirming the predictive power of behavioral economics indicators in markets. Notably, the BERT-based Daily Sentiment Index (DSI) exhibits a 1- to 3-day lead time in predicting market turning points, offering a novel tool for real-time market monitoring.

### 6.2 Research Contributions

The innovative contributions of this study can be categorized into three main areas: theoretical, methodological, and empirical. Theoretically, this research establishes a multidimensional behavioral analysis framework by integrating high-frequency trading data, social media text, and institutional holdings. This approach overcomes the limitations of traditional methods that rely on a single data source, allowing for a more comprehensive understanding of investor behavior. Methodologically, the study introduces a novel LSTM-ATT prediction model, which combines deep learning with attention mechanisms. This model retains the advantages of time-series modeling while enhancing the identification of critical nodes, improving the accuracy of behavioral predictions. Empirically, the research applies structural equation modeling to quantify the transmission pathways of different behavioral biases. The findings reveal significant differences in the effects of overconfidence and herd behavior on market volatility, with path coefficients of 0.72 and 0.58, respectively.

The study also has several practical implications for different market participants. For regulators, a multi-source investor behavior monitoring system is recommended, which can trigger risk alerts when abnormal order ratios exceed predefined thresholds, such as the mean plus two standard deviations. Institutional investors can benefit from sentiment divergence indices to construct hedging strategies, with empirical results demonstrating a Sharpe ratio of 1.35 in out-of-sample tests. For individual investors, the study highlights the importance of recognizing disposition effects in their trading behavior. Implementing systematic stop-loss disciplines can help mitigate losses and improve long-term investment outcomes. Overall, the findings contribute to both academic research and market practice, offering valuable insights for regulators, institutional investors, and individual market participants alike.

### 6.3 Limitations

This study has several limitations that require further improvement. First, the scope of the data is primarily limited to China's A-share market. While the findings provide valuable insights into investor behavior within this market, additional validation is necessary to determine their applicability across different financial markets. Future research should extend the dataset to include global markets, allowing for a more comprehensive understanding of behavioral biases in diverse regulatory environments. Second, despite utilizing advanced BERT models for text analysis, challenges related to semantic ambiguity and sarcasm detection remain. These are common issues in natural language processing (NLP) that can affect the accuracy of sentiment analysis. Further advancements in NLP techniques, including the integration of contextual learning and multimodal data, may help improve the reliability of text-based sentiment assessments. Third, the study assumes that investor behavior patterns remain relatively stable over time. However, structural changes in the financial markets, such as the increasing prevalence of algorithmic trading, may significantly alter behavioral biases. Future research should incorporate dynamic modeling approaches to account for evolving market conditions and investor strategies.

Given these limitations, several future research directions can be explored. One promising avenue is conducting cross-market comparative studies to examine how behavioral biases manifest differently under varying regulatory regimes. Understanding these differences can provide valuable insights for policymakers and investors in different financial systems. Another direction involves developing multimodal sentiment analysis frameworks that incorporate visual data, such as financial videos and graphical content, to enhance sentiment prediction accuracy. This approach can complement traditional text-based sentiment analysis by capturing additional emotional and contextual cues. Additionally, with the rapid development of blockchain technology and decentralized finance (DeFi), future studies can investigate new behavioral characteristics in blockchain-based financial environments. Decentralized decision-making processes in DeFi platforms may introduce unique behavioral patterns that differ from those observed in traditional financial markets. Exploring these aspects can contribute to a deeper understanding of investor behavior in emerging financial ecosystems.

### 6.4 Implications for Future Financial Market Research

Based on the findings of this study, several policy recommendations are proposed to enhance market stability and investor protection. First, regulatory authorities should promote regulatory technology (RegTech) innovation by developing an "Investor Behavior Dashboard" system. This system would integrate key behavioral indicators such as abnormal order ratios and sentiment divergence, allowing for real-time visual monitoring of investor activity. By leveraging advanced tools like natural language processing (NLP) for social media sentiment analysis and big data analytics for detecting anomalous trading patterns, regulators can improve market surveillance and risk management. The implementation of such technologies can help identify market anomalies early and prevent excessive volatility caused by behavioral biases.

Second, regulatory measures should be refined to address the behavioral traits of different investor groups. Targeted supervision can be implemented to mitigate risks associated with specific biases. For instance, dynamic margin requirements could be introduced for accounts that exhibit overconfidence, such as those with a monthly turnover exceeding 300%. Institutional investors prone to herd behavior should be subject to enhanced position disclosure requirements, improving market transparency. Additionally, retail investors who are significantly affected by loss aversion should receive risk warnings and protective measures to prevent excessive losses driven by emotional decision-making.

Finally, optimizing market infrastructure is crucial in reducing the impact of behavioral biases on trading outcomes. Exchanges could introduce "cooling-off" mechanisms to counteract the disposition effect, such as delaying the execution of large loss-making trades to give investors time for reassessment. Furthermore, investor education programs should incorporate principles of behavioral economics to help individuals recognize and correct cognitive biases. By providing investors with a deeper understanding of their own behavioral tendencies, financial literacy programs can contribute to more rational decision-making and improved long-term investment performance. These policy measures, when implemented collectively, can foster a more stable and efficient financial market.

**Reference**

[1] Tetlock, P. C., Saar-Tsechansky, M., & Macskassy, S. A. (2008). More than words: Quantifying language to measure firms' fundamentals. *Journal of Finance, 63*(3), 1437-1467. https://doi.org/10.1111/j.1540-6261.2008.01367.x

[2] Shiller, R. J. (2000). *Irrational exuberance*. Princeton University Press.

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

[4] Barberis, N., Shleifer, A., & Vishny, R. W. (1998). A model of investor sentiment. *Journal of Financial Economics, 49*(3), 307-343. https://doi.org/10.1016/S0304-405X(98)00027-0

[5] Lo, A. W. (2004). The adaptive markets hypothesis: Market efficiency from an evolutionary perspective. *Journal of Portfolio Management, 30*(5), 15-29. https://doi.org/10.3905/jpm.2004.442721

[6] Graham, J. R., & Harvey, C. R. (2001). The theory and practice of corporate finance: Evidence from the field.

*Journal of Financial Economics, 60*(2-3), 187-243. https://doi.org/10.1016/S0304-405X(01)00044-7

[7] Barber, B. M., & Odean, T. (2001). Boys will be boys: Gender, overconfidence, and common stock investment. *The Quarterly Journal of Economics, 116*(1), 261-292.

[8] Bholat, D. (2015). Big data and central banks. *Bank of England Quarterly Bulletin, 55*(2), 92-101.

[9] Bollen, J., Mao, H., & Zeng, X. (2011). Twitter mood predicts the stock market. *Journal of Computational Science, 2*(1), 1-8.

[10] Chen, Z., Zhang, W., & Huang, Y. (2022). The integration of behavioral finance and big data analytics in stock market prediction. *Journal of Financial Data Science, 4*(2), 55-70.

[11] Feng, G., Giglio, S., & Xiu, D. (2019). Taming the factor zoo: A test of new factors. *Journal of Finance, 74*(5), 2209-2244.

[12] Gao, X., Han, Y., & Li, J. (2020). Machine learning applications in behavioral finance: A review. *Finance Research Letters, 35*, 101310.

[13] Glasserman, P., & Kang, C. (2018). Design of stress scenarios using empirical likelihood. *Management Science, 64*(11), 5100-5119.

[14] Kahneman, D., & Tversky, A. (1979). Prospect theory: An analysis of decision under risk. *Econometrica, 47*(2), 263-291.

[15] Kahneman, D., & Tversky, A. (1992). Advances in prospect theory: Cumulative representation of uncertainty. *Journal of Risk and Uncertainty, 5*(4), 297-323.

[16] Luo, Y., Wu, C., & Zhang, X. (2021). The role of big data in identifying financial market anomalies. *Journal of Financial Markets, 54*, 101429.

[17] Preis, T., Moat, H. S., & Stanley, H. E. (2013). Quantifying trading behavior in financial markets using Google Trends. *Scientific Reports, 3*, 1684.

[18] Shiller, R. J. (2003). From efficient markets theory to behavioral finance. *Journal of Economic Perspectives, 17*(1), 83-104.

[19] Tetlock, P. C. (2007). Giving content to investor sentiment: The role of media in the stock market. *Journal of Finance, 62*(3), 1139-1168.

[20] Tversky, A., & Kahneman, D. (1974). Judgment under uncertainty: Heuristics and biases. *Science, 185*(4157), 1124-1131.

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

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

[23] De Bondt, W. F. M., & Thaler, R. (1985). Does the stock market overreact? *Journal of Finance*, 40(3), 793-805. https://doi.org/10.1111/j.1540-6261.1985.tb05004.x

[24] Shiller, R. J. (2003). From efficient markets theory to behavioral finance. *Journal of Economic Perspectives*, 17(1), 83-104. https://doi.org/10.1257/089533003321164967

[25] Fama, E. F. (1998). Market efficiency, long-term returns, and behavioral finance. *Journal of Financial Economics*, 49(3), 283-306. https://doi.org/10.1016/S0304-405X(98)00026-9

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

[27] Ben-David, I., Franzoni, F., & Moussawi, R. (2018). Do ETFs increase volatility? *Journal of Finance*, 73(6), 2471-2535. https://doi.org/10.1111/jofi.12727

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

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

[30] Da, Z., Engelberg, J., & Gao, P. (2015). The sum of all FEARS investor sentiment and asset prices. *Review of Financial Studies*, 28(1), 1-32. https://doi.org/10.1093/rfs/hhu072

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

[32] Hong, H., & Stein, J. C. (2007). Disagreement and the stock market. *Journal of Economic Perspectives*, 21(2), 109-128. https://doi.org/10.1257/jep.21.2.109

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

[34] Li, F., Lundholm, R., & Minnis, M. (2013). A measure of competition based on 10-K filings. *Journal of Accounting Research*, 51(2), 399-436. https://doi.org/10.1111/1475-679X.12010

[35] Tetlock, P. C., Saar-Tsechansky, M., & Macskassy, S. (2008). More than words: Quantifying language to measure firms' fundamentals. *Journal of Finance*, 63(3), 1437-1467. https://doi.org/10.1111/j.1540-6261.2008.01362.x

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

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

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

[39] Devlin, J., Chang, M. W., Lee, K., & Toutanova, K. (2019). BERT: Pre-training of deep bidirectional transformers for language understanding. *Proceedings of NAACL-HLT 2019*, 4171-4186. https://doi.org/10.18653/v1/N19-1423

[40] Chen, T., & Guestrin, C. (2016). XGBoost: A scalable tree boosting system. *Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*, 785-794. https://doi.org/10.1145/2939672.2939785

[41] Bollerslev, T. (1986). Generalized autoregressive conditional heteroskedasticity. *Journal of Econometrics*, 31(3), 307-327. https://doi.org/10.1016/0304-4076(86)90063-1

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

[43] Shiller, R. J. (2003). From efficient markets theory to behavioral finance. *Journal of Economic Perspectives*, 17(1), 83-104. https://doi.org/10.1257/089533003321164967

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

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