Data Science and Engineering (DSE) Record, Volume 6, issue 1

ANALYSIS OF HERDING BEHAVIOR IN STOCK MARKETS USING CROSS-SECTIONAL ABSOLUTE DEVIATION AND GATED RECURRENT UNIT

[1]Thanaphat Sriboonma and [2]Sumalee Sangamuang

[1]Master's Degree Program in Data Science, Chiang Mai University, Chiang Mai, Thailand

[2]College of Art, Media and Technology, Chiang Mai University, Chiang Mai, Thailand
[1]thanaphat_sriboonma@cmu.ac.th, [2]sumalee.sa@cmu.ac.th

Abstract

Herding behavior, where investors mimic the trading actions of others, is a critical phenomenon in financial markets, often exacerbating volatility and leading to mispricing and systemic risk. Traditional econometric models like the Cross-Sectional Absolute Deviation (CSAD) have been widely used to detect such behavior, yet they fall short in capturing the nonlinear, dynamic nature of market sentiment. This study integrates behavioral finance with deep learning to enhance the prediction of herding behavior by estimating the CSAD-based herding coefficient (γ_2) using advanced time-series models. Daily stock data from the S&P 500,spanning January 2000 to December 2024, is analyzed using four deep learning architectures: Recurrent Neural Network (RNN), Gated Recurrent Unit (GRU), Long Short-Term Memory (LSTM), and Time-Series BERT (TST-BERT). The models are evaluated using Mean Squared Error (MSE), Root Mean Squared Error (RMSE), Mean Absolute Error (MAE), and the Coefficient of Determination (R^2). Among them, the GRU model outperformed the others, achieving the lowest prediction error and highest R^2 value of 0.8785, indicating its superior capability in modeling temporal dependencies in financial data. The results affirm that deep learning, particularly GRU, provides a more accurate and robust framework for detecting herding behavior, offering valuable insights for investors, regulators, and policymakers aiming to enhance market stability and risk assessment.

Keyword: Herding Behavior, Cross-sectional Absolute Deviation, Deep Learning, Regression, Recurrent Neural Network, Gated Recurrent Unit, Long Short Term Memory, Time-Series BER

1. Introduction

Herding behavior in financial markets is a well-documented phenomenon in which investors tend to imitate the trading decisions of others rather than relying on their own independent analysis. This behavior can lead to asset mispricing, increased market volatility, and systemic risk, potentially destabilizing financial markets[1]. Herding is often categorized into two types: informational herding, where investors follow the consensus due to perceived superior information, and behavioral herding, where irrational factors such as fear, overconfidence, or speculative motives drive collective decision-making[2][3]. During periods of market stress, such as the Global Financial Crisis (GFC) of 2008 and the COVID-19 pandemic, herding tendencies are amplified as uncertainty leads to mass panic selling or speculative buying, further exacerbating market instability[4].

The understanding of Herding behavior in the context of the stock market is crucial because it sheds light on why markets can sometimes exhibit extreme volatility, speculative bubbles and unexpected price movements[5]. When herding is widespread, individual stock prices may move in highly correlated patterns, reducing the benefits of diversification and making markets more susceptible to systemic shocks[6].

Among the most commonly employed methodologies for detecting herding behavior in financial markets is the Cross-Sectional Absolute Deviation (CSAD), which measures the dispersion of individual stock returns relative to the overall market return[7][8]. The non-linear CSAD model suggests that when investors suppress independent decision-making and follow market trends collectively, the dispersion of individual stock returns tends to decrease as market movements intensify, signaling herding behavior[8]. Empirical studies have applied CSAD across both developed and emerging markets, finding that herding effects are more pronounced during periods of heightened market uncertainty and financial crises[6].

Although various econometric models and statistical techniques have been widely employed to analyze herding behavior in financial markets, these traditional approaches face several critical limitations. Specifically, they struggle to capture the complex, non-linear relationships inherent in financial time-series data and often fail to adapt to the time-varying nature and high-frequency fluctuations characteristic of large-scale financial datasets. These shortcomings become particularly evident during periods of heightened market volatility or financial crises, where investor sentiment shifts rapidly and herding behavior intensifies. As a result, traditional models tend to produce suboptimal forecasts, limiting their effectiveness in guiding investment strategies and risk management[8]. This represents a significant research gap, highlighting the need for more adaptive and robust modeling techniques. Leveraging advanced machine learning and deep learning methods offers a promising avenue to overcome these limitations by providing greater flexibility, improved accuracy, and the ability to model dynamic interactions across numerous financial instruments simultaneously.

With the advancement of machine learning and deep learning techniques, researchers have increasingly recognized the limitations of traditional econometric models in capturing complex, non-linear relationships inherent in financial time-series data. Traditional econometric models, such as linear regression and factor-based methods, typically assume linearity and stationary conditions, making them insufficient in modeling the dynamic, evolving nature of investor behavior, particularly herding behavior observed in volatile market conditions.

To overcome these limitations, researchers have adopted more advanced neural network models. Recurrent Neural Networks (RNNs) have emerged as an initial improvement over traditional models due to their capability to handle sequential data effectively. Studies have demonstrated that even simple RNN architectures outperform conventional econometric approaches in forecasting financial time-series data[9]. However, RNNs are susceptible to the vanishing gradient problem, significantly hindering their ability to retain long-term dependencies, which are critical in accurately predicting stock market trends and investor behavior over extended periods[10].

Addressing the shortcomings of standard RNNs, advanced architectures such as Gated Recurrent Units (GRUs) and Long Short-Term Memory Networks (LSTMs) have been developed. These models incorporate gating mechanisms that selectively retain relevant historical information while discarding irrelevant or redundant data, effectively mitigating the vanishing gradient issue[11]. Empirical research highlights the superior performance of GRUs and particularly LSTMs, in forecasting financial anomalies such as extreme market movements, asset mispricing, and market sentiment influenced by investor biases[12][13].

Most recently, transformer-based architectures like Time-Series BERT (TST-BERT) have introduced a significant advancement in financial time-series analysis. TST-BERT employs self-attention mechanisms capable of modeling long-range dependencies and complex interactions across multiple financial instruments simultaneously. This ability to handle parallel sequences and interdependencies makes TST-BERT especially suited for large-scale financial datasets, such as the S&P 500, where comprehensive analysis of stock interactions and macroeconomic indicators is crucial. Nonetheless, despite its promising capabilities, TST-BERT faces challenges in computational efficiency, as its self-attention mechanism requires substantial computational resources, particularly with lengthy input sequences[12].

2. Literature Review

2.1 Definition and Theoretical Background of Herding Behavior

Herding behavior in financial markets refers to the tendency of investors to mimic the trading actions of others, often driven by factors such as informational asymmetry, reputation concerns, psychological biases, and market uncertainty[1]. This behavior can be categorized as either rational or irrational herding.

The rational perspective suggests that investors follow the crowd due to strategic incentives or informational cascades. Scharfstein and Stein[2] argue that fund managers imitate others to protect their reputations, while Trueman[14] and Graham[15] suggest that financial analysts often issue forecasts that align with previous predictions to avoid reputational risk. In a similar manner, Froot et al.[16] explain that short-term speculators herd to extract private information from other investors. The concept of informational cascades, introduced by Bikhchandani et al.[17], describes how investors may disregard their private signals and rely on the observed actions of others, leading to asset bubbles and market distortions[5].

The behavioral finance perspective views herding as the result of psychological and social influences rather than rational investment strategies. Keynes[18] suggests that during periods of uncertainty, investors follow the crowd due to social conventions. Shleifer and Summers[19] differentiate between rational arbitrageurs and noise traders, arguing that irrational investors often drive herding behavior based on sentiment rather than fundamentals. Barberis et al.[20] propose that investor sentiment can lead to market overreaction and underreaction, while Daniel et al.[21] suggest that overconfidence bias amplifies herding tendencies.

2.2 Empirical Studies on Herding Behavior

Empirical research on herding behavior primarily follows two approaches: institutional investor herding and market-wide herding. Studies on institutional herding focus on how large market participants influence asset prices. Lakonishok et al.[22] and Grinblatt et al.[23] find limited herding among US pension and mutual fund managers but note stronger herding in small-cap stocks. Wermers[24] also finds much higher levels of herding in trading by growth-oriented mutual funds. Sias[25] and Choi and Sias[26] further confirm that institutional investors tend to follow one another's trades over time, increasing correlation in asset prices.

Studies on market-wide herding provide mixed results. Christie and Huang[7] employ the Cross-Sectional Standard Deviation (CSSD) approach and find no significant evidence of herding in US markets. Chang et al.[8] introduce the Cross-Sectional Absolute Deviation (CSAD) method, which reduces sensitivity to outliers, and report significant herding in South Korea and Taiwan, with weaker evidence in Japan. Chiang and Zheng (2010) examine 18 markets and find strong herding effects in Asian and emerging markets but little evidence in the US or Latin America. Economou et al.[6] identify cross-market herding effects in Southern Europe, particularly during economic crises.

2.3 Measuring Herding Behavior: CSSD vs. CSAD

Several methodologies have been developed to quantify herding behavior, with the CSSD and CSAD models being the most widely used.

Cross-Sectional Standard Deviation (CSSD)

Christie and Huang[7] introduce the CSSD method to measure the dispersion of individual stock returns from the market return. The formula is given by:

$$CSSD_t = \sqrt{\frac{\sum_{i=1}^{N}(R_{i,t} - R_{m,t})^2}{N-1}}$$

where:

- $CSSD_t$ represents the cross-sectional standard deviation of returns at time t,
- $R_{i,t}$ is the return of stock i at time t,
- $R_{m,t}$ is the market return at time t,
- N is the number of stocks in the sample.

A lower CSSD value suggests stronger herding behavior, as stock returns converge toward the market return. However, CSSD is highly sensitive to extreme market movements, making it less robust in volatile environments.

Cross-Sectional Absolute Deviation (CSAD)

To address CSSD's limitations, Chang et al.[8] propose the CSAD measure, which uses absolute deviations instead of squared deviations. The formula for CSAD is:

$$CSAD_t = \frac{1}{N}\sum_{i=1}^{N}|R_{i,t} - R_{m,t}|$$

where:

- $CSAD_t$ represents the cross-sectional absolute deviation at time t,
- $|R_{i,t}-R_{m,t}|$ measures the absolute deviation of stock iii from the market return.

Chang et al.[8] also introduce a non-linear regression model to test for herding:

$$CSAD_t = \gamma_0 + \gamma_1|R_{m,t}| + \gamma_2 R_{m,t}^2 + \varepsilon_t$$

where:

- γ_2 is the key coefficient. If γ_2 is significantly negative, it indicates herding behavior, as stock return dispersion decreases during large market movements.

The CSAD approach is more robust than CSSD and has been widely used in recent studies [6]

2.4 Herding Behavior and Market Conditions

Herding behavior varies across different market conditions and is often more pronounced during periods of financial turbulence and heightened volatility. Klein[27] suggests that, in times of financial distress, investors are more likely to follow collective market trends rather than rely on fundamental analysis. Supporting this view, Chiang et al.[28] show that herding intensified during the later stages of the Asian financial crisis, as stock return correlations remained persistently high.

However, herding behavior is not uniform and can exhibit asymmetry. Hwang and Salmon[29] introduce the concept of beta herding—a non-parametric measure based on the cross-sectional variation in market betas. Their findings indicate that investors tend to align their trading behaviors with the perceived market direction, particularly during stable periods of market growth or decline. Interestingly, they observe that herding weakens during crises, as investors tend to shift their focus toward fundamental valuations rather than collective sentiment.

2.5 Advances in Herding Prediction: Machine Learning Approaches

The prediction of herding behavior in financial markets has traditionally relied on econometric models, such as Ordinary Least Squares (OLS) regression, Autoregressive Integrated Moving Average (ARIMA), and Generalized Autoregressive Conditional Heteroskedasticity (GARCH) models[30][31]. These statistical techniques have been widely used for analyzing financial data and capturing volatility dynamics[32]. However, traditional econometric models often rely on assumptions such as linearity and stationarity, which may not hold in real-world financial data. These models can struggle to capture complex, non-linear dependencies, sudden regime shifts, and high-frequency market interactions, which are crucial for understanding investor sentiment and trading behavior. For instance, high-frequency financial data often exhibit nonstationarity and low signal-to-noise ratios, posing challenges for traditional analysis methods[33]. Additionally, the assumption of linearity in models like Vector Autoregression (VAR) may limit their applicability in capturing the intricate patterns present in high-frequency trading data[34]. Moreover, traditional models may not effectively process non-traditional datasets, such as textual data from social media, which are increasingly important for gauging investor sentiment[35].

With the increasing popularity of computational intelligence in finance, fueled by rapid advancements in artificial intelligence (AI) and deep learning (DL), machine learning (ML) techniques have gained significant traction in financial forecasting. These methods have demonstrated strong predictive capabilities and adaptability to evolving market conditions[36]. One of the most promising machine learning techniques for time-series forecasting is the Recurrent Neural Network (RNN)[37]. Studies have found that even simple neural network architectures can be effective and often outperform traditional econometric models[9]. However, a significant limitation of standard RNNs is the vanishing gradient problem, which hinders their ability to capture long-term dependencies in financial time-series data[10][38].

To address this issue, Gated Recurrent Units (GRUs) and Long Short-Term Memory (LSTM) networks have been developed as improved versions of RNNs[10][11]. These models incorporate gating mechanisms that selectively retain and discard information, improving their ability to learn long-range dependencies in financial data[39]. Studies have shown that LSTMs outperform traditional econometric models in predicting stock market movements and capturing investor sentiment[40]. GRUs outperform LSTM networks on low complexity sequences while on high complexity sequences the order is reversed[41].

More recently, transformer-based architectures such as Time-Series BERT (TST-BERT) have emerged as a powerful alternative to RNN-based models. Originally introduced for Natural Language Processing (NLP), the Bidirectional Encoder Representations from Transformers (BERT) model has been adapted for time-series forecasting, demonstrating state-of-the-art performance in financial prediction tasks[12]. Unlike RNNs and LSTMs, TST-BERT employs a self-attention mechanism, allowing it to capture complex relationships, long-term dependencies, and dynamic interactions across multiple financial variables[42]. The application of the Transformer also faces some challenges. First, the high computational resource requirements of the Transformer model limit its application in resource-constrained environments. Second, the lack of interpretability poses a risk when the model is applied in security-sensitive financial scenarios, as users may find it difficult to understand the model's decision-making process[43].

Several studies have explored the application of deep learning models in detecting and predicting trends in financial markets. For instance, Wu et al.[44] explored the application of various deep learning models, including LSTM variants, CNNs, and Transformer models, for cryptocurrency price forecasting. Their study focused on evaluating the performance of these models under different market conditions, particularly during the COVID-19 pandemic. The results indicated that univariate LSTM model variants outperformed other architectures in predicting cryptocurrency prices, demonstrating the models' robustness in handling high market volatility. Additionally, their analysis highlighted significant price fluctuations across major cryptocurrencies, underscoring the importance of selecting appropriate deep learning models for financial forecasting tasks.

Similarly, in their study, Gao and Zhang[45] developed a composite investor sentiment index based on six indicators—five objective and one subjective—to investigate the impact of investor sentiment on price volatility in China's capital market from a behavioral finance perspective. They employed a hybrid neural network model combining Variational Mode Decomposition (VMD) and Long Short-Term Memory (LSTM) to decompose both the investor sentiment index and the Shanghai Stock Exchange Composite Index (SSEC) into short-term, medium-term, and long-term trends. Each component was separately trained to generate forecasts across different time scales, which were then aggregated to produce the final prediction. Their results showed that this hybrid model significantly improved forecasting accuracy compared to traditional approaches. Additionally, they used GARCH and co-integration error regression models to examine fluctuation correlations, and VAR (Vector Auto-Regression) models to analyze causal relationships between investor sentiment and stock index movements.

Furthermore, Su et al.[46] conducted a comprehensive review of transformer-based architectures for long-term time series forecasting (LTSF). Their study outlined the evolution of transformer models and their adaptations for LTSF tasks, highlighting their ability to capture semantic correlations and long-range dependencies more efficiently than traditional RNN-based models such as LSTM and GRU. Although transformers offer computational advantages and modeling strength, they also face challenges related to time complexity and potential loss of interdependencies during optimization. The authors discussed best practices for training transformers, provided an overview of available LTSF datasets and evaluation metrics, and proposed future research directions, including the integration of compound models and large language models (LLMs) to enhance the accuracy, interpretability, and transparency of forecasting outcomes. These advancements position transformer-based models as a powerful and evolving tool for financial time series analysis and prediction.

2.6 Conclusion of Literature Review

Recent studies on herding behavior mostly use traditional econometric models, such as CSSD and CSAD, along with other factors to improve the performance of predicting herding behavior. This research has evolved from traditional econometric models, as shown in Table 1, to machine learning-based prediction approaches. Deep learning techniques, particularly RNN, GRU, LSTM, and TST-BERT, offer new opportunities to improve herding detection and forecasting by identifying hidden patterns in financial data.

Table 1: Literature of studies focusing on herding behavior

Year	Authors	Title	Tool	Dataset
2024	Destan Kirimhan , James E. Payne , Osamah AlKhazali [47]	Herd behavior in U.S. bank stocks	-CASD -stock market sentiment -turnover of trading volume in the market -TMP period	Daily stock data of the U.S. banks covered by CRSP for the trading days between June 2, 2014 and December 30, 2022
2024	Natividad Blasco , Luis Casas , Sandra Ferreruela [48]	Does war spread the herding effect in stock markets? Evidence from emerging and developed markets during the Russia-Ukraine war	-CASD	The 23 countries belonging to the MSCI World Index and the 24 countries included in the MSCI Emerging Markets Index and Russia. between January 2021 and February 2023.

Table 1 (continued)

Year	Authors	Title	Tool	Dataset
2024	Xolani Sibande [49]	Herding behaviour and monetary policy: Evidence from the ZAR market	-CASD - the policy rate -the policy rate announcements	The South African currency market. The daily currency market data ranges from January 1, 2000 to March 3, 2022
2023	Iftekhar Hasan , Radu Tunaru , Davide Vioto [50]	Herding behavior and systemic risk in global stock markets	-CASD -A five-factor asset pricing model -TED spread	Asia-Pacific,Latin American,North American, European markets in period from January 1 2000 to December 31 2022.
2023	Huu Manh Nguyen, Walid Bakry , Thi Huong Giang Vuong [51]	COVID-19 pandemic and herd behavior: Evidence from a frontier market	-CASD -Quantile regression (QR)	The stocks listed on the HOSE and the HNX from January 2016 to May 2022.
2022	Mouna Youssef , Sami Sobhi Waked [52]	Herding behavior in the cryptocurrency market during COVID-19 pandemic: The role of media coverage	-CASD -a dummy variable that takes one during the COVID-19 crisis	The top-43 cryptocurrencies in terms of market capitalization from April 28 2013 to November 11 2020.

This research aims to bridge behavioral finance and AI-driven analysis, enhancing the understanding of herding behavior in financial markets. By leveraging advanced deep learning models to predict herding in S&P 500 stocks, this study contributes to better market stability, investment decision-making, and risk assessment.

3. Dataset and Methodology

Dataset

This study utilizes daily stock price data (High, Low, and Close) of the S&P 500 index stock over the period from January 1, 2000, to December 31, 2024, obtained from Yahoo Finance. On each trading day, the dataset records the price of the S&P 500 index, forming a time-series structure.

Methodology

Step 1: Data Preparation and CSAD Construction

This stage involves the collection, preprocessing, and transformation of financial market data required for herding behavior analysis. For each trading day, the return of the S&P 500 index stock is computed and used to represent the overall market behavior. The degree of price fluctuation over time is quantified using the CSAD model, which measures the deviation of the index return from its average market trend.

The CSAD at time t is defined as

$$\mathrm{CSAD}_t = \frac{1}{N_t}\sum_{i=1}^{N_t} |R_{i,t} - R_{m,t}|,$$

where $R_{i,t}$ denotes the return of stock i and $R_{m,t}$ represents the market return at time. This formulation provides a daily measure of cross-sectional dispersion that serves as a preliminary indicator of collective movement in the market.

For example, consider a single stock on a trading day, with a stock return of -0.706% and a market return of -0.388%

The CSAD for that day is calculated as

$$\mathrm{CSAD_t} = \frac{|(-0.706)-(-0.388)|}{1} = 0.318$$

A higher CSAD value indicates greater dispersion among individual stock returns, implying weaker herding, whereas a lower CSAD suggests stronger movement and possible herding behavior.

The resulting daily CSAD series constitutes the foundation for estimating the herding coefficients (γ_0, γ_1, γ_2) in Step 2.

Step 2: Herding Model Estimation (Baseline Regression)

Building on the CSAD series constructed in Step 1, the herding coefficients are estimated from the canonical CSAD regression:

$$\mathbf{CSAD}_t = \gamma_0 + \gamma_1 |R_{m,t}| + \gamma_2 R_{m,t}^2 + \varepsilon_t,$$

Where $R_{m,t}$ is the market return at time . The coefficient γ_2 captures the nonlinear sensitivity of cross-sectional dispersion to market movements; a significantly negative γ_2 is consistent with herding.

$$\gamma_2 = \frac{\mathrm{N}\times\Sigma(CSAD\times R_m^2)-(\Sigma\mathrm{CASD}\times\Sigma(R_m^2))}{\mathrm{N}\times\Sigma(R_m^4)-(\Sigma(\mathrm{R}_m^2))^2} = \frac{(20\times2.085)-(7.904\times6.197)}{(20\times12.047)-(6.197)^2} = -0.036 \text{ (Herding Behavior)}$$

Step 3: Deep Learning Models for Herding Coefficient Prediction

To model the temporal evolution of herding behavior, four deep learning architectures tailored to sequential time-series data are employed: Recurrent Neural Network (RNN), Gated Recurrent Unit (GRU), Long Short-Term Memory (LSTM), and a lightweight Transformer encoder (referred to here as TRANS). Each model learns from fixed-length input sequences of lagged behavioral and market features to predict the herding coefficient $\gamma_{2,t}$ constructed in Step 2.

The RNN provides a simple recurrent baseline. GRU introduces gating to regulate information flow and improve optimization stability. LSTM further enhances memory retention via distinct input/forget/output gates, capturing short-term dynamics more robustly. The Transformer uses self-attention to model both local and longer-range dependencies, offering a non-recurrent alternative.

Preprocessing and fairness controls. All predictors are lagged by one period to prevent look-ahead. Features are standardized with a scale fitted on training data only and then applied to validation/test segments. Models are trained and evaluated under the same K-Fold Cross-Validation splits to ensure a fair comparison.

Input–output structure. For each fold, we create sequences of length SEQ_WIN so that the model sees the most recent history (t−SEQ_WIN,…,t−1) and predicts $\gamma_{2,t}$. No shuffling is used at any stage.

Full code listings can be placed in the colab; this section focuses on the design choices and shared training protocol.

Table 2: Deep learning model configurations for herding coefficient prediction

Model	Core Layer/ Block	Hidden Size/ d_model	Loss	Optimizer	Max Epochs	SEQ_ WIN	Notes
RNN	Simple RNN	layer 1:256 layer 2:256 dropout1:0.2 dropout2:0.4	MSE	Adam (Learning Rate = 0.00315)	150	1	Tuned using Keras Tuner
GRU	GRU	layer 1:128 layer 2:352 dropout1:0.1 dropout2:0.1	MSE	Adam (Learning Rate = 0.00676)	170	1	Tuned using Keras Tuner

Table 2 (continued)

Model	Core Layer/ Block	Hidden Size/ d_model	Loss	Optimizer	Max Epochs	SEQ_ WIN	Notes
LSTM	LSTM	layer:64 layer 2:192 dropout1:0.1 dropout2:0.1	MSE	Adam (Learning Rate = 0.00676)	180	1	Tuned using Keras Tuner
TST-BERT	Trans former Encoder	d_model:192 num_heads:8 num_layer:3 dropout:0.1	MSE	Adam (Learning Rate = 5.98e-05)	200	1	Tuned using Optuna

Shared training protocol (applies to all models):

- Evaluation method: K-Fold Cross-Validation with identical splits across models (no shuffling).
- Feature handling: all features shifted by one period; standardization fit on train only.
- Batch size: modest (e.g., 8) to stabilize training on small samples.

Step 4: Hyperparameter Optimization

To further improve the model's predictive performance, hyperparameter tuning is conducted. For the RNN, GRU, and LSTM models, Keras Tuner is used to identify the optimal set of hyperparameters such as learning rate, batch size, and number of layers. For the TST-BERT model, Optuna is employed due to its efficiency in navigating complex search spaces. This tuning process ensures that each model operates under the most favorable conditions for learning and generalization.

Step 5: K-Fold Cross-Validation

To ensure a robust evaluation of the predictive models, a K-Fold Cross-Validation procedure is implemented. No shuffling is applied in order to maintain the chronological order of observations and prevent look-ahead bias, which is important when modeling sequential financial data.

The dataset is divided into K sequential folds (here, K = 5). For each fold:

- The training set includes all observations up to a certain point in time, while the validation set contains the subsequent observations.
- The model is trained on the training set and evaluated on the validation set using a consistent set of performance metrics.
- Model predictions are rescaled to the original range of the herding coefficient γ_2 before computing evaluation metrics, ensuring interpretability.
- This process is repeated for all K folds, providing an overall assessment of model generalization performance.

The performance metrics used in each fold include Mean Squared Error (MSE), Root Mean Squared Error (RMSE), Mean Absolute Error (MAE), and the Coefficient of Determination (R^2). After completing all folds, the mean of these metrics are reported to quantify model reliability and robustness.

This K-Fold Cross-Validation procedure ensures a fair comparison among models and provides a solid foundation for selecting the final model for herding coefficient prediction.

Step 6: Model Evaluation and Selection

Each model is evaluated based on its predictive accuracy, with performance measured using four key metrics: Mean Squared Error (MSE), Root Mean Squared Error (RMSE), Mean Absolute Error (MAE), and the Coefficient of Determination (R^2). The GRU model serves as a reference benchmark. The model that achieves the best performance compared to GRU across these metrics is selected for the final estimation of γ_2.

Step 7: Interpretation and Detection of Herding Behavior

After estimating the herding coefficient (γ_2) using the CSAD model and its deep learning extensions, the final step involves interpreting the results to determine the presence and intensity of herding behavior in the market. In the canonical CSAD regression, a negative and statistically significant coefficient of the squared market return term (γ_2) indicates the existence of herding behavior. This suggests that as market returns deviate from the mean, individual asset returns tend to converge toward the overall market return rather than diverge, reflecting collective investor movement.

In the deep learning framework, herding behavior is detected through the predicted herding coefficient values and the error metrics (MSE, RMSE, MAE, and R^2). Consistently accurate predictions of γ_2 with low error rates imply that the model effectively captures nonlinear and time-dependent patterns associated with investor imitation. By comparing the predicted and actual γ_2 values across time, periods of intensified herding can be identified—typically during times of heightened volatility, market stress, or major economic events.

4. Results

This study investigates the effectiveness of four deep learning models—Recurrent Neural Network (RNN), Gated Recurrent Unit (GRU), Long Short-Term Memory (LSTM), and Time-Series BERT (TST-BERT)—in forecasting the herding behavior coefficient (γ_2), which is derived from the Cross-Sectional Absolute Deviation (CSAD) model. The primary goal is to evaluate the models' ability to detect herding behavior in financial markets using standard regression performance metrics, including Mean Squared Error (MSE), Root Mean Squared Error (RMSE), Mean Absolute Error (MAE), and the Coefficient of Determination (R^2).

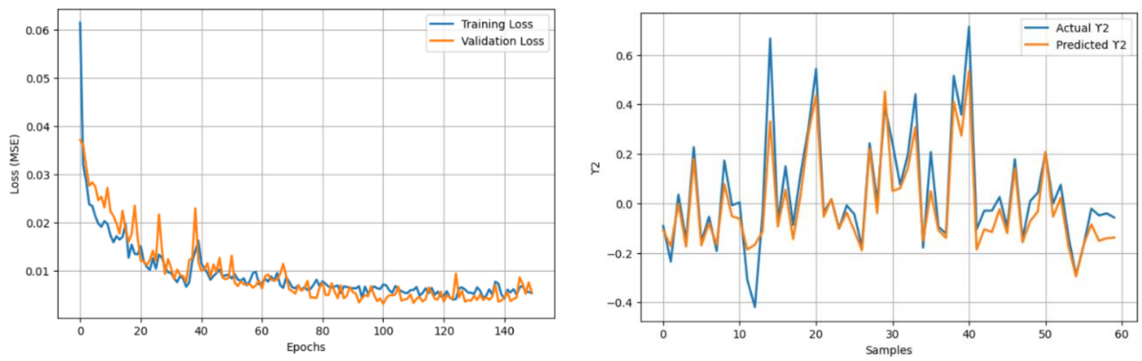
The performance of the Recurrent Neural Network (RNN) model in predicting the herding coefficient was evaluated using 5-fold cross-validation. Table 4 summarizes the results across all folds in terms of Mean Squared Error (MSE), Root Mean Squared Error (RMSE), Mean Absolute Error (MAE), and the coefficient of determination (R^2).

Overall, the model demonstrated stable and consistent performance across all folds, with an average MSE of 0.0064, RMSE of 0.0790, MAE of 0.0563, and an average R^2 of 0.8350. These results indicate that the RNN effectively captured the temporal dependencies within the dataset and achieved a high level of predictive accuracy.

Figure 1(a) illustrates the training and validation loss curves over 150 epochs. Both curves show a consistent decline and convergence, suggesting that the model successfully learned the underlying patterns without overfitting. Figure 1(b) compares the actual and predicted values of for the test set, showing that the predicted values closely follow the actual trend, confirming the model's ability to generalize well to unseen data.

Table 3: Performance of RNN across 5-fold cross-validation

Fold	MSE	RMSE	MAE	R^2
Fold 1	0.0060	0.0773	0.0522	0.8847
Fold 2	0.0069	0.0831	0.0636	0.8606
Fold 3	0.0102	0.1009	0.0694	0.7302
Fold 4	0.0037	0.0606	0.0476	0.8802
Fold 5	0.0053	0.0731	0.0488	0.8196
Average	0.0064	0.0790	0.0563	0.8350



(a) Training loss and Validation loss (b) Actual values and Predicted values

Figure 1. Graph of RNN

The performance of the Gated Recurrent Unit (GRU) model in predicting the herding coefficient was evaluated using 5-fold cross-validation. Table 5 presents the performance metrics across all folds, including Mean Squared Error (MSE), Root Mean Squared Error (RMSE), Mean Absolute Error (MAE), and the coefficient of determination (R^2).

The GRU model achieved an average MSE of 0.0048, RMSE of 0.0685, MAE of 0.0474, and an average R^2 of 0.8785. These results indicate that the GRU effectively captured the sequential dependencies in the data while maintaining a strong generalization capability.

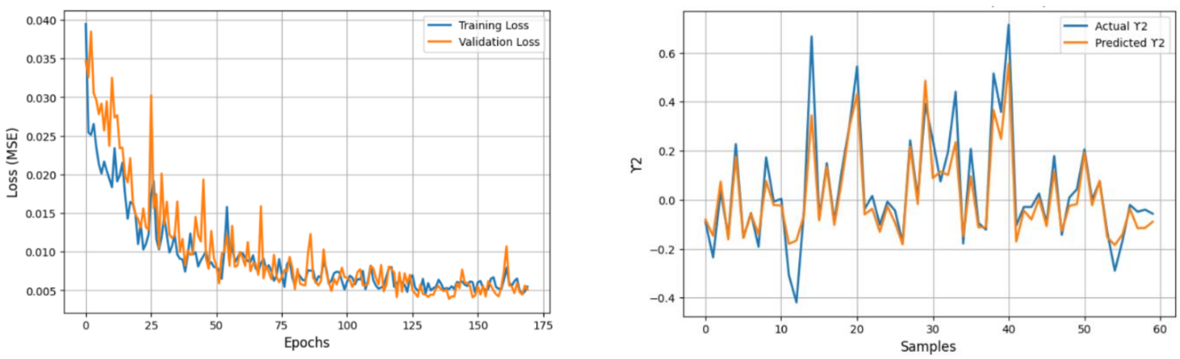
Figure 2(a) shows the training and validation loss curves over 170 epochs. Both losses decreased steadily and converged smoothly, implying that the GRU model learned effectively without significant overfitting. Figure 2(b) compares the actual and predicted values of , where the predicted trend closely follows the actual observations, confirming the model's reliable predictive ability.

Table 4: Performance of GRU across 5-fold cross-validation

Fold	MSE	RMSE	MAE	R^2
Fold 1	0.0048	0.0691	0.0480	0.9079
Fold 2	0.0063	0.0794	0.0603	0.8728
Fold 3	0.0050	0.0706	0.0444	0.8680
Fold 4	0.0030	0.0546	0.0413	0.9028

Table 4 (continued)

Fold	MSE	RMSE	MAE	R^2
Fold 5	0.0047	0.0686	0.0431	0.8411
Average	0.0048	0.0685	0.0474	0.8785



(a) Training loss and Validation loss (b) Actual values and Predicted values

Figure 2. Graph of GRU

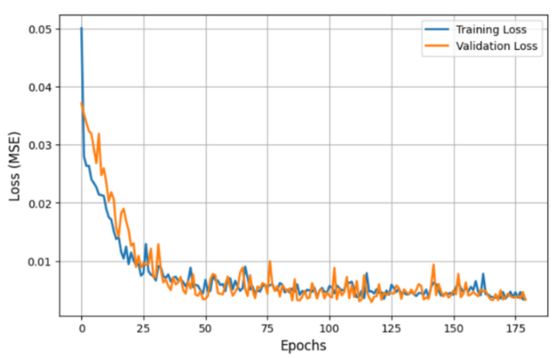
The Long Short-Term Memory (LSTM) model was implemented to further enhance the ability to capture long-term dependencies in sequential data for herding coefficient prediction. The model's performance was evaluated using 5-fold cross-validation, and the results are summarized in Table 6, which presents the Mean Squared Error (MSE), Root Mean Squared Error (RMSE), Mean Absolute Error (MAE), and the coefficient of determination (R^2) for each fold.

The LSTM model achieved an average MSE of 0.0053, RMSE of 0.0727, MAE of 0.0517, and an average R^2 of 0.8604. These results show that the LSTM maintained high predictive accuracy and performed comparably to the GRU model, with slightly better stability across folds. The performance demonstrates that the LSTM effectively learned both short-term and long-term patterns within the time-series data.

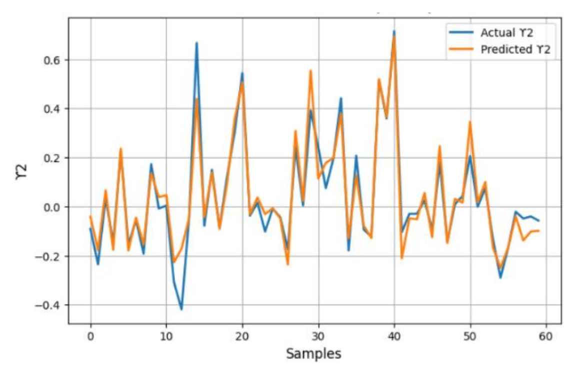
Figure 3(a) illustrates the training and validation loss curves over 180 epochs. Both losses decreased steadily and converged at a low value, indicating effective learning and minimal overfitting. Figure 3(b) shows the comparison between actual and predicted values of , where the predicted line closely follows the actual trend, confirming the LSTM's ability to generalize to unseen data.

Table 5: Performance of LSTM across 5-fold cross-validation

Fold	MSE	RMSE	MAE	R^2
Fold 1	0.0044	0.0662	0.0459	0.9154
Fold 2	0.0065	0.0808	0.0598	0.8683
Fold 3	0.0060	0.0775	0.0513	0.8409
Fold 4	0.0042	0.0648	0.0507	0.8631
Fold 5	0.0055	0.0742	0.0509	0.8140
Average	0.0053	0.0727	0.0517	0.8604



(a) Training loss and Validation loss



(b) Actual values and Predicted values

Figure 3. Graph of LSTM

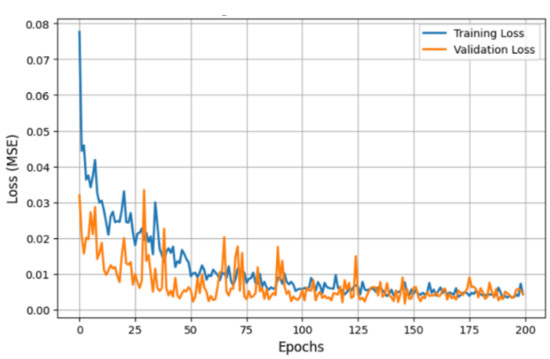
The Time-Series Transformer based on BERT (TST-BERT) model was employed to leverage the attention mechanism for learning complex temporal dependencies in herding coefficient prediction. The model's performance was evaluated using 5-fold cross-validation, and the detailed results are summarized in Table 7, which presents the Mean Squared Error (MSE), Root Mean Squared Error (RMSE), Mean Absolute Error (MAE), and the coefficient of determination (R^2) across all folds.

The TST-BERT model achieved an average MSE of 0.0168, RMSE of 0.1208, MAE of 0.0932, and an average R^2 of 0.6550. While the model exhibited relatively higher errors compared to RNN-based models such as GRU and LSTM, it still demonstrated the capability to capture the overall pattern and fluctuations of the herding coefficient. The variation in R^2 scores across folds suggests that model performance may be sensitive to specific training subsets, likely due to the model's higher complexity and data dependency.

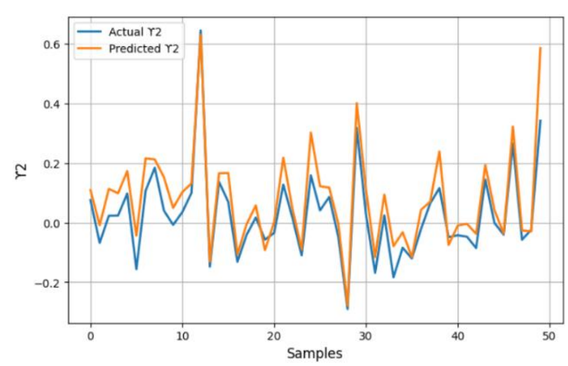
Figure 4(a) illustrates the training and validation loss curves across 200 epochs, showing that both losses decreased and stabilized over time, indicating effective convergence. Figure 4(b) compares the actual and predicted values of , showing that the predicted series successfully followed the general trend of the actual values, though with minor deviations at certain peaks.

Table 6: Performance of TST-BERT across 5-fold cross-validation

Fold	MSE	RMSE	MAE	R^2
Fold 1	0.0406	0.2016	0.1587	0.4627
Fold 2	0.0089	0.0944	0.0737	0.6381
Fold 3	0.0160	0.1264	0.1011	0.7036
Fold 4	0.0146	0.1207	0.0862	0.6352
Fold 5	0.0037	0.0612	0.0464	0.8357
Average	0.0168	0.1208	0.0932	0.6550



(a) Training loss and Validation loss



(b) Actual values and Predicted values

Figure 4. Graph of TST-BERT

A summary of the models' performance across the four evaluation metrics is presented in Table 8. Among the four models, the Gated Recurrent Unit (GRU) achieved the best overall performance, with the lowest Mean Squared Error (MSE = 0.0048), Root Mean Squared Error (RMSE = 0.0685), Mean Absolute Error (MAE = 0.0474), and the highest coefficient of determination (R^2 = 0.8785). This comparison reinforces the conclusion that the GRU model offers the most accurate and robust performance in predicting the herding coefficient. Its ability to maintain low error rates and achieve high predictive alignment confirms its suitability for financial time series forecasting in this context.

In conclusion, the GRU model emerges as the most effective tool for identifying herding behavior, as indicated by its superior predictive accuracy and model stability. The final interpretation of herding dynamics in the market is derived from the γ_2 values predicted by the GRU model. A significantly negative γ_2 indicates the presence of herding among investors during the observed period, providing valuable insights into market sentiment and behavioral finance.

Table 7: The performance of each model

Performance	Mean Squared Error (MSE)	Root Mean Squared Error (RMSE)	Mean Absolute Error (MAE)	The Coefficient of Determination (R^2)
Recurrent Neural Network (RNN)	0.0064	0.0790	0.0563	0.8350
Gated Recurrent Unit (GRU)	0.0048	0.0685	0.0474	0.8785

Table 7 (continued)

Performance	Mean Squared Error (MSE)	Root Mean Squared Error (RMSE)	Mean Absolute Error (MAE)	The Coefficient of Determination (R^2)
Long Short Term Memory (LSTM)	0.0053	0.0727	0.0517	0.8604
Time-Series BERT (TST-BERT)	0.0168	0.1208	0.0932	0.6550

5. Discussion

This study set out to improve the detection and prediction of herding behavior in financial markets by applying the Cross-Sectional Absolute Deviation (CSAD) method with advanced deep learning models. The main objective was to estimate the herding coefficient (γ_2)—a central indicator of collective investor behavior—using time-series approaches based on four model types: RNN, GRU, LSTM, and TST-BERT. The results provide several important insights.

Among the models, the Gated Recurrent Unit (GRU) gave the most accurate forecasts, achieving the lowest prediction errors (MSE 0.0048, RMSE 0.0685, MAE 0.0474) and the highest explanatory power ($R^2 = 0.8785$). Its ability to track patterns in financial time series more effectively than the standard RNN or LSTM reflects its efficiency in retaining useful historical information while filtering out noise. This makes the GRU particularly well-suited to capturing the dynamics of investor behavior.

Earlier work (e.g., Chung et al. [11]) has shown that recurrent models with gating mechanisms, such as GRU and LSTM, perform better than traditional RNNs, which often fail to capture long-term influences in market data. Consistent with this, the GRU in this study outperformed, while the RNN performed fairly well but not as effectively as GRU and LSTM (MSE 0.0064, RMSE 0.0790, MAE 0.0563, R^2 = 0.8350), reflecting its limited ability to capture deeper structures in investor behavior.

The LSTM model produced reasonably strong results but performed slightly weaker than the GRU (MSE 0.0046, RMSE 0.0682, MAE 0.0472, R^2 = 0.7232), consistent with findings from Ghadimpour and Ebrahimi [53]. Although LSTMs are designed to capture long-term temporal dependencies through separate memory and gating mechanisms, their more complex architecture increases computational burden, slows convergence, and makes them more sensitive to hyperparameter tuning. These limitations became evident in the current dataset, where LSTM required more careful parameter adjustments yet still did not match the performance of GRU.

Additional empirical evidence reinforces the advantages of GRU over LSTM. Zarzycki and Ławryńczuk [55] found that GRU networks can approximate dynamical processes with accuracy comparable to LSTMs while requiring significantly fewer parameters. In their predictive control experiments for two chemical reactors, GRU achieved similar predictive performance but with greater computational efficiency, leading the authors to recommend GRU for real-time modeling in model predictive control (MPC). Likewise, comparative research on EEG-based prediction tasks (Rivas et al. [56]) reported that while LSTMs possess strong memory capacity, GRUs offer faster training, higher efficiency, and better adaptability when datasets are limited or vary in size. Together, these studies align with the present findings by demonstrating that the simpler GRU architecture often delivers equal—or superior—performance with lower computational costs, making it more practical and robust in many time-series forecasting contexts.

By contrast, the transformer-based TST-BERT model yielded the weakest performance (MSE 0.0168, RMSE 0.1208, MAE 0.0932, $R^2 = 0.6550$). While transformer architectures are theoretically well-suited for capturing long-term dependencies in sequential data, they generally demand larger datasets and greater computational resources to achieve stable results. In this study, the relatively limited dataset contributed to fluctuating validation losses and unstable predictions, consistent with the observations of Su et al. [46], who reported similar challenges when applying transformers to small-scale financial data.

The results not only highlight the suitability of GRU for this type of analysis but also confirm the value of the CSAD framework. Using γ_2 as a target variable within modern forecasting techniques provides a way to dynamically track herding tendencies. Importantly, the GRU model indicated periods of significantly negative γ_2 values in the S&P 500 between 2000 and 2024, pointing to herding behavior during times of crisis such as the COVID-19 pandemic, geopolitical tensions, and heightened market volatility.

In sum, this study shows that combining behavioral finance tools with modern forecasting methods can enrich the analysis of market psychology. It suggests that future work could benefit from broader datasets, the inclusion of additional factors such as investor sentiment or macroeconomic indicators, and the exploration of hybrid models that balance interpretability with predictive power.

6. Conclusion

This study aimed to bridge the gap between behavioral finance and artificial intelligence by predicting herding behavior in the stock market using the CSAD methodology enhanced through deep learning models. By focusing on the S&P 500 dataset from 2000 to 2024, and estimating the herding coefficient γ_2, this research offers an advanced, data-driven approach for understanding collective investor behavior under uncertain market conditions.

Among the four deep learning architectures evaluated—Recurrent Neural Network (RNN), Gated Recurrent Unit (GRU), Long Short-Term Memory (LSTM), and Time-Series BERT (TST-BERT)—the GRU model demonstrated superior performance across all evaluation metrics. It achieved the lowest Mean Squared Error (0.0048), the lowest RMSE (0.0685), the lowest MAE (0.0474), and the highest R^2 value (0.8785), indicating robust predictive accuracy and a strong fit to actual herding behavior patterns. The LSTM model followed closely, showing strong results in handling long-term dependencies but slightly underperforming the GRU in overall error metrics, with an MSE of 0.0053, RMSE of 0.0727, MAE of 0.0517, and R^2 of 0.8604. The RNN model showed moderate success, achieving an MSE of 0.0064, RMSE of 0.0790, MAE of 0.0563, and R^2 of 0.8350, but its limitations due to the vanishing gradient problem hindered its ability to fully capture complex temporal patterns. TST-BERT, despite its theoretical advantages in modeling long-range dependencies, underperformed in this context with an MSE of 0.0168, RMSE of 0.1208, MAE of 0.0932, and R^2 of 0.6550—likely due to the complexity of the model relative to the dataset size and its sensitivity to training dynamics.

The findings confirm that GRU is the most reliable model for forecasting the herding coefficient γ_2 in financial time-series data. A significantly negative γ_2 value predicted by GRU indicates strong evidence of market-wide herding behavior, particularly under volatile market conditions. Ultimately, this research not only reaffirms the value of the CSAD model for detecting herding but also underscores the power of deep learning—particularly GRU-based architectures—in financial forecasting. The integration of behavioral finance insights with AI-driven methodologies opens new avenues for improving market stability, optimizing investment decisions, and enhancing systemic risk assessment in increasingly dynamic global markets.

Moreover, this study opens several avenues for future research, particularly in extending herding behavior forecasting models across different market structures, such as cross-country or cross-industry analyses. Incorporating macroeconomic variables and financial uncertainty indices into deep learning frameworks could further enhance the models' ability to capture market dynamics during periods of financial stress. In addition, integrating Explainable Artificial Intelligence (XAI) techniques with the GRU model may improve the transparency and interpretability of the forecasting process, thereby increasing the model's credibility in the context of policy formulation and systemic risk management. Overall, the integration of behavioral finance and deep learning proposed in this study not only contributes to the existing literature but also offers substantial potential for practical applications in investment decision-making and the promotion of long-term financial market stability.

References

[1] Sushil Bikhchandani and Sunil Sharma, "Herd behavior in financial markets", IMF Staff Papers, vol. 47, no. 3, pp. 279-310, 2001.

[2] David S. Scharfstein and Jeremy C. Stein, "Herd Behavior and Investment", The American Economic Review, vol. 80, no. 3, pp. 465-479, 1990.

[3] Soosung Hwang and Mark Salmon , "Market stress and herding" , Journal of Empirical Finance, vol. 11, no. 4, pp. 585-616, 2004.

[4] Emilios Galariotis , Wu Rong and Spyros I. Spyrou , "Herding on fundamental information: A comparative study" , Journal of Banking & Finance, vol. 50, pp. 589-598, 2015.

[5] Isik Akin and Meryem Akin , " Behavioral finance impacts on US stock market volatility: an analysis of market anomalies" , Behavioural Public Policy, https://doi.org/10.1017/bpp.2024.13, 2024.

[6] Fotini Economou , Alexandros Kostakis and Nikolaos Philippas , "Cross-country effects in herding behaviour: Evidence from four south European markets" , Journal of International Financial Markets, Institutions and Money, vol. 21, no. 3, pp. 443-460, 2011.

[7] William G. Christie and Roger D. Huang, "Following the Pied Piper: Do Individual Returns Herd around the Market?", Financial Analysts Journal, vol. 51, no. 4, pp. 31-37, 1995.

[8] Eric C Chang, Joseph W Cheng and Ajay Khorana , "An examination of herd behavior in equity markets: An international perspective" , Journal of Banking & Finance, vol. 24, no. 10, pp. 1651-16790, 2000.

[9] Eero Tölö, " Predicting systemic financial crises with recurrent neural networks", Journal of Financial Stability, vol. 49, no. 100746, 2020.

[10] Sepp Hochreiter and Jürgen Schmidhuber, "Long Short-Term Memory", Neural Computation, vol. 9, no. 8, pp. 1735-1780, 1997.

[11] Junyoung Chung, Caglar Gulcehre, KyungHyun Cho and Yoshua Bengio, "Empirical Evaluation of Gated Recurrent Neural Networks on Sequence Modeling", arXiv preprint arXiv:1412.3555,2014.

[12] Cheng Zhang, Nilam Nur Amir Sjarif and Roslina Ibrahim, "Deep learning models for price forecasting of financial time series: A review of recent advancements: 2020-2022", arXiv preprint arXiv:2305.04811,2023

[13] Lucian Liviu Albu and Radu Lupu, "Anomaly detection in stock market indices with neural networks" , Journal of Financial Studies, vol. 5, no. 9, pp. 10-23, 2000.

[14] Brett Trueman, "Analyst Forecasts and Herding Behavior", The Review of Financial Studies, vol. 7, no. 1, pp. 97-124, 1994.

[15] John R. Graham, "Herding among Investment Newsletters: Theory and Evidence", The Journal of Finance, vol. 54, no. 1, pp. 237-268, 1999

[16] Kenneth A. Froot, David S. Scharfstein and Jeremy C. Stein, "Herd on the Street : Informational Inefficiencies in a Market with Short-Term Speculation", The Journal of Finance, vol. 47, no. 4, pp. 1461-1484, 1992.

[17] Sushil Bikhchandani, David Hirshleifer and Ivo Welch, "A Theory of Fads, Fashion, Custom, and Cultural Change as Informational Cascades" , Journal of Political Economy, vol. 100, no. 5, pp. 992-1026, 1992.

[18] John Maynard Keynes, "The General Theory of Employment, Interest, and Money", Macmillan, 1936.

[19] Andrei Shleifer and Lawrence H. Summers, "The Noise Trader Approach to Finance", The Quarterly Journal of Economics, vol. 4, no. 2, pp. 19-33, 1990.

[20] Barberis Nicholas, Andrei Shleifer, and Robert Vishny, "A Model of Investor Sentiment" , Journal of Financial Economics, vol. 49, no. 3, pp. 307-343, 1998.

[21] Kent Daniel, David Hirshleifer and Avanidhar Subrahmanyam, "Investor Psychology and Security Market under- and Overreactions", The Journal of Finance, vol. 53, no. 6, pp. 1839-1885, 1998.

[22] Josef Lakonishok, Andrei Shleifer and Robert W. Vishny, "The impact of institutional trading on stock prices" , Journal of Financial Economics, vol. 32, no. 1, pp. 23-43, 1992.

[23] Mark Grinblatt, Sheridan Titman and Russ Wermers, "Momentum Investment Strategies, Portfolio Performance, and Herding: A Study of Mutual Fund Behavior", The American Economic Review, vol. 85, no. 5, pp. 1088-1105, 1995.

[24] Russ Wermers, "Mutual Fund Herding and the Impact on Stock Prices", The Journal of Finance, vol. 54, no. 2, pp. 581-622, 1999.

[25] Richard W. Sias, "Institutional Herding", The Review of Financial Studies, vol. 17, no. 1, pp. 165-206, 2004.

[26] Nicole Y. Choi and Richard W. Sias, "Institutional Industry Herding" , Journal of Financial Economics, vol. 94, no. 3, pp. 469-491, 2009.

[27] Arne C. Klein, "Time-variations in herding behavior: Evidence from a Markov switching SUR model" , Journal of International Financial Markets, Institutions and Money, vol. 26, pp. 291-304, 2013.

[28] Thomas C. Chiang , Bang Nam Jeon and Huimin Li , "Dynamic correlation analysis of financial contagion: Evidence from Asian markets" , Journal of International Money and Finance, vol. 26, no. 7, pp. 1206-1228 , 2007.

[29] Soosung Hwang and Mark Salmon, " Sentiment and Beta Herding", SSRN Electronic Journal, 2006.

[30] Robert S. Pindyck and Daniel L. Rubinfeld, " Econometric Models and Economic Forecasts (4th ed.)", McGraw-Hill/Irwin, 1998.

[31] Robert F. Engle, "Autoregressive Conditional Heteroscedasticity with Estimates of the Variance of United Kingdom Inflation", The Review of Financial Studies, vol. 50, no. 4, pp. 987-1007, 1982.

[32] Ruey S. Tsay, "Analysis of Financial Time Series (3rd ed.)", Wiley, 2010.

[33] Lu Zhang and Lei Hua, "Major Issues in High-Frequency Financial Data Analysis: A Survey of Solutions", Mathematics, vol. 13, no. 3, pp. 347, 2025.

[34] Abdul Rahman and Neelesh Upadhye, "Hybrid Vector Auto Regression and Neural Network Model for Order Flow Imbalance Prediction in High Frequency Trading", arXiv preprint arXiv:2411.08382,2024.

[35] Ajit Desai, "Machine Learning for Economics Research: When What and How?", arXiv preprint arXiv:2304.00086,2023.

[36] Ahmet Murat Ozbayoglu, Mehmet Ugur Gudelek and Omer Berat Sezer, "Deep learning for financial applications : A survey", Applied Soft Computing Journal, , vol. 93, no. 106384, 2020.

[37] Hansika Hewamalage, Christoph Bergmeir, Kasun Bandara, "Recurrent Neural Networks for Time Series Forecasting: Current status and future directions", arXiv preprint arXiv:1909.00590,2020.

[38] Yoshua Bengio, Patrice Y. Simard and P. Frasconi, "Learning long-term dependencies with gradient descent is difficult" , IEEE Transactions on Neural Networks and Learning Systems, vol. 5, no.2 , pp. 157-166, 1994.

[39] Thomas Fischer and Christopher Krauss, "Deep learning with long short-term memory networks for financial market predictions" , European Journal of Operational Research, vol. 270, no.2 , pp. 654-669, 2018.

[40] Meet Satishbhai Sonani, Atta Badii and Armin Moin, "Stock Price Prediction Using a Hybrid LSTM-GNN Model: Integrating Time-Series and Graph-Based Analysis", arXiv preprint arXiv:2502.15813, 2025.

[41] Roberto Cahuantzi, Xinye Chen and Stefan Güttel, "A comparison of LSTM and GRU networks for learning symbolic sequences", arXiv preprint arXiv:2107.02248, 2023.

[42] Bryan Lim, Sercan O. Arik, Nicolas Loeff and Tomas Pfister, "Temporal Fusion Transformers for Interpretable Multi-horizon Time Series Forecasting", arXiv preprint arXiv:1912.09363, 2020.

[43] Tianxu Liu, Yanbin Wang, Jianguo Sun, Ye Tian, Yanyu Huang, Tao Xue, Peiyue Li and Yiwei Liu, "The Role of Transformer Models in Advancing Blockchain Technology: A Systematic Survey", arXiv preprint arXiv:2409.02139, 2024.

[44] Jingyang Wu, Xinyi Zhang, Fangyixuan Huang, Haochen Zhou and Rohtiash Chandra, "Review of deep learning models for crypto price prediction: implementation and evaluation", arXiv preprint arXiv:2405.11431, 2024.

[45] Zhenbin Gao and Jie Zhang, "The fluctuation correlation between investor sentiment and stock index using VMD-LSTM: Evidence from China stock market" , The North American Journal of Economics and Finance, vol. 66, no. 101915 , 2023.

[46] Liyilei Su, Xumin Zuo, Rui Li, Xin Wang, Heng Zhao and Bingding Huang, " A Systematic Review for Transformer-based Long-term Series Forecasting", arXiv preprint arXiv:2310.20218, 2023.

[47] Destan Kirimhan, James E. Payne and Osamah AlKhazali, "Herd behavior in U.S. bank stocks", Finance Research Letters, vol. 67, Part B, no. 105930, 2024.

[48] Natividad Blasco, Luis Casas and Sandra Ferreruela, "Does war spread the herding effect in stock markets? Evidence from emerging and developed markets during the Russia-Ukraine war", Finance Research Letters, vol. 63, no. 105365, 2024.

[49] Xolani Sibande, "Herding behaviour and monetary policy: Evidence from the ZAR market" , Journal of Behavioral and Experimental Finance, vol. 42, no. 100920, 2024.

[50] Iftekhar Hasan, Radu Tunaru and Davide Vioto , "Herding behavior and systemic risk in global stock markets" , Journal of Empirical Finance, vol. 73, no. 10, pp. 107-133, 2023.

[51] Huu Manh Nguyen, Walid Bakry and Thi Huong Giang Vuong, "COVID-19 pandemic and herd behavior: Evidence from a frontier market" , Journal of Behavioral and Experimental Finance, vol. 38, no. 100807, 2023.

[52] Mouna Youssef and Sami Sobhi Waked, "Herding behavior in the cryptocurrency market during COVID-19 pandemic: The role of media coverage" ,The North American Journal of Economics and Finance, vol. 62, no. 101752, 2022.

[53] Mohammadreza Ghadimpour and Seyed babak Ebrahimi, "Forecasting Financial Time Series Using Deep Learning Networks: Evidence from Long-Short Term Memory and Gated Recurrent Unit", Iranian Journal of Finance, vol. 6, no. 4, pp. 81-94, 2022.

[54] Mohammadreza Ghadimpour and Seyed babak Ebrahimi, "Forecasting Financial Time Series Using Deep Learning Networks: Evidence from Long-Short Term Memory and Gated Recurrent Unit", Iranian Journal of Finance, vol. 6, no. 4, pp. 81-94, 2022.

[55] Krzysztof Zarzycki and Maciej Ławryńczuk, "LSTM and GRU Neural Networks as Models of Dynamical Processes Used in Predictive Control: A Comparison of Models Developed for Two Chemical Reactors", Sensors, vol. 21, no. 16, Art. no. 5625, 2021.

[56] Fernando Rivas, Jesús Enrique Sierra-Garcia and Jose María Camara, "Comparison of LSTM- and GRU-Type RNN Networks for Attention and Meditation Prediction on Raw EEG Data from Low-Cost Headsets", Electronics, vol. 14, Art. no. 707, 2025.