Mathematics

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

AIMS Mathematics, 11(1): 977–1020.
DOI: 10.3934/math.2026043
Received: 21 October 2025
Revised: 31 December 2025
Accepted: 08 January 2026
Published: 13 January 2026

Research article

A hybrid transformer framework integrating sentiment and dynamic market structure for stock price movement forecasting

Dong-Jun Kim[1], **Eunjung Noh**[2] **and Sun-Yong Choi**[1,2,*]

[1] Department of Finance and Big data, Gachon University, Seongnam 13120, Republic of Korea
[2] Department of Mathematics and Computer Science, Rollins College, Winter Park, Florida, USA

* **Correspondence:** Email: sunyongchoi@gachon.ac.kr; Tel: +82317505387.

Abstract: Traditional financial forecasting models are inherently limited as their sole reliance on price data causes them to overlook critical information such as market sentiment and dynamic inter-market interactions. To address this shortcoming, this study proposes a novel hybrid transformer model that integrates heterogeneous data sources. Specifically, our framework combines investor sentiment extracted from news headlines using FinBERT and dynamic changes in market structure analyzed with a TVP-VAR model, along with traditional log-return data. In forecasting experiments conducted on four major global stock indices (S&P 500, FTSE 100, CSI 300, and Nikkei 225), the proposed model consistently demonstrated superior performance compared to both single-data-source models and traditional benchmarks. We found that the inclusion of dynamic market structure information, derived from the TVP-VAR model, as a predictive variable was a decisive factor in improving forecasting accuracy. This research empirically validates that a multi-modal approach combining heterogeneous data can significantly enhance the precision of financial market forecasting.

Keywords: transformer; TVP-VAR; FinBERT; multi-modal; hybrid-transformer-model; stock price
Mathematics Subject Classification: 91G70, 91B84, 62M10, 68T07, 68T09, 91G80

1. Introduction

Financial markets constitute a dynamic system driven by multilayered factors such as global capital flows, policy shifts, and corporate earnings. While the log return serves as a key metric for volatility measurement and risk management, accurate forecasting remains highly challenging due to the inherent non-stationarity, volatility clustering, and unpredictable exogenous shocks embedded in financial data [1, 2]. Furthermore, existing price-based models often fail to fully capture the complex dynamics of the market. This failure stems primarily from their inability to integrate and analyze non-price latent factors, such as investor psychological responses (sentiment) and the dynamically evolving structure of inter-country market interactions. The omission of this critical information is the core research problem limiting predictive accuracy.

To overcome these fundamental limitations and maximize forecasting precision, this study proposes a novel hybrid Transformer framework that systematically integrates heterogeneous data sources—text (sentiment), structure (spillover effects), and price (log returns)—representing our primary contribution. Specifically, we employ FinBERT to extract quantitative sentiment variables from news headlines and utilize the Time-Varying Parameter Vector Autoregression (TVP-VAR) model to quantify the dynamic spillover effects among national indices. By integrating these multi-layered exogenous features into a Transformer-based deep learning model to learn long-range dependencies, we empirically demonstrate a significant enhancement in predictive performance unattainable by conventional single-data-source models. This integrated methodology offers substantial practical value for portfolio optimization, algorithmic trading, and systemic risk mitigation.

To highlight the methodological novelty of our approach, we summarized comparative studies from 2021 to 2025 in Table 1. As shown in the table, existing literature can be broadly categorized into two streams. Group A focuses on monitoring dynamic market structures and systemic risk using models like TVP-VAR but often overlooks the unstructured sentiment information embedded in news text [3–5]. Conversely, Group B leverages advanced NLP models like FinBERT to capture investor sentiment but typically restricts analysis to individual assets or fails to explicitly model the time-varying spillover effects among markets [6–8]. In contrast, our study distinguishes itself by organically integrating both micro-level investor sentiment (via FinBERT) and macro-level dynamic market structure (via TVP-VAR) within a unified hybrid Transformer framework, thereby overcoming the limitations of these isolated approaches.

Recent deep learning approaches have extended beyond the short-term dependency learning of Recurrent Neural Network (RNN) variants (Long Short-Term Memory (LSTM), Gated Recurrent Unit (GRU)), drawing attention to the ability of Transformer architectures to capture long-range dependencies [9, 10]. Thanks to the self-attention mechanism, Transformers can learn interactions across entire input sequences in parallel, effectively alleviating the vanishing gradient problem that intensifies with sequence length. When pre-trained language models (e.g., Bidirectional Encoder Representations from Transformers (BERT), Financial BERT (FinBERT)) originally developed for natural language processing are adapted to financial time series, they enable multi-modal learning across unstructured text and structured price data, yielding significant improvements in prediction accuracy and training stability compared to conventional RNNs [11, 12]. To exploit these structural advantages, this study integrates Convolutional Neural Network (CNN)-based feature extractors and an LSTM-based preprocessing module within a unified Transformer framework.

Price movements in financial markets often cannot be fully explained by historical price data alone, as investors' psychological responses tend to precede volatility shifts. Consequently, there has been a surge of interest in quantifying market sentiment from text sources - such as news articles, official announcements, and social media - and using it as an auxiliary predictor [13, 14]. Domain-specific pre-trained language models offer superior understanding of financial terminology and context, enabling fine-grained interpretation of tone and nuance beyond simple positive, negative, neutral classification [15, 16]. In this study, we incorporate country-specific sentiment indices into the Transformer input to capture spillover effects that price-based models alone may miss, thereby further enhancing forecasting performance.

Table 1. Comparison of recent related works (2021–2025) and the novelty of the proposed framework. Existing studies typically focus on either market structure (Group A) or sentiment analysis (Group B) in isolation.

Reference	Methodology	Key Characteristics & Limitations
Group A: Focus on Dynamic Market Structure		
[3]	TVP-VAR + Deep Learning	Monitors systematic financial risk using dynamic connectedness;
		Lacks textual sentiment integration.
[4]	TVP-VAR + Quantile VAR	Analyzes dynamic return and volatility spillovers across multi-asset markets;
		Relies solely on market data, neglecting unstructured sentiment inputs.
[5]	TVP-VAR + Deep Learning	Forecasts exchange rate spillovers using dynamic market structure;
		Does not account for unstructured investor sentiment data.
Group B: Focus on Investor Sentiment		
[6]	FinBERT + LSTM	Forecasts S&P 500 movements using mathematically derived sentiment analysis;
		Relies on LSTM, lacking dynamic market structure analysis.
[7]	FinBERT + Deep Learning	Predicts bank stock prices using sentiment analysis;
		Focuses on individual assets without inter-market spillover effects.
[8]	FinBERT + Numerical Masking	Enhances sentiment accuracy via numerical change-related masking;
		Focuses on textual understanding without integrating dynamic market structure.
Proposed Framework		
This Study	TVP-VAR + FinBERT	Integrates both dynamic market structure (macro) and
	+ Hybrid Transformer	investor sentiment (micro) into a unified Transformer model.

Building on the potential of alternative data sources, this study moves beyond simple variable addition to rigorously examine the efficacy of systematic information integration. Specifically, to validate the superiority of the proposed approach, we formulate the following research question and measurable hypothesis

- Research Question: Does a multi-modal deep learning framework that integrates unstructured text data (sentiment) and structural market data (spillover effects) with historical prices significantly outperform traditional single-source models in forecasting financial time series?
- Hypothesis: We hypothesize that the proposed hybrid Transformer model, which incorporates FinBERT-derived sentiment indices and TVP-VAR-based spillover features, will yield statistically lower prediction errors (e.g., Root Mean Square Error (RMSE), Mean Absolute Error (MAE)) compared to baseline models relying solely on historical log returns or single-modal inputs.

Accordingly, the core objective of this study is to improve prediction accuracy by integrating latent factors that are difficult to capture with price-based models alone. To this end, we first convert investment sentiment, which varies with news, into a variable through the sentiment analysis of news headlines. Second, we explicitly model the dynamic spillover effects among different stock indices. By applying these multi-layered exogenous variables to an integrated deep learning framework, we aim to learn the complex market interactions with greater precision and maximize predictive performance, thereby demonstrating the practical value of the proposed model.

In this study, we propose a hybrid Transformer-based framework to capture the complex characteristics inherent in financial time series data. The proposed architecture consists of three core modules, each specializing in processing unstructured text information, dynamic inter-market interactions, and inherent time-series patterns.

First, we utilize FinBERT, a language model specialized for the financial domain, to extract market sentiment from news headline data [15, 17]. This process transforms the dynamic changes in investor sentiment, which are difficult to ascertain from price data alone, into quantitative features that are used as input variables for the model [18, 19].

Second, we adopt the TVP-VAR model to explicitly capture the time-varying spillover effects between countries [20, 21]. Because interactions among assets in countries are constantly changing in response to policy shifts and macroeconomic shocks, the TVP-VAR model can effectively estimate these dynamic relationships [22, 23].

Third, we use a Transformer as the core prediction model to learn the long-term dependencies inherent in time series data [9, 24, 25]. Unlike traditional RNN-based models, the Transformer's self-attention mechanism directly computes the relationships between all points within a sequence [26, 27]. This enables the Transformer model to select crucial past information and dynamically incorporate its influence into predictions, thereby effectively learning complex and non-linear patterns [28, 29].

Accordingly, the proposed model is an organic hybrid architecture that extracts multi-faceted features through specialized modules tailored to each data type's characteristics and integrates them into a Transformer to learn long-term time-series patterns. Through this approach, we aim to maximize the predictive performance for the log returns of each country's stock index by comprehensively analyzing multivariate financial data.

Through the proposal and empirical validation of this integrated framework, this study makes two primary academic contributions.

First, it offers a methodological contribution by proposing a novel hybrid deep learning architecture that systematically integrates multi-source data. Unlike existing research that relies on single sources of information, our model quantifies market participant sentiment through FinBERT and the dynamic changes in market structure through a combination of TVP-VAR and CNN. In other words, the framework proposed in this study integrally considers both the psychological factors of investors, quantified through FinBERT, and the structural factors of the market's endogenous interconnectedness, estimated by the TVP-VAR model. By fusing qualitative, text-based information reflecting investor sentiment with quantitative information capturing the time-varying market structure through TVP-VAR, it becomes possible to analyze the dynamics of individual assets comprehensively within the context of their interaction with the entire market system, thereby moving beyond a fragmented perspective. Conclusively, this integrated methodology, centered on the dynamic structural analysis enabled by TVP-VAR, overcomes the limitations in the explanatory power of single-data-source models and makes a core contribution to the significant enhancement of predictive performance by more precisely capturing the complex and non-linear dynamics of financial markets.

Second, this study empirically demonstrates that the interconnectedness between markets is a variable that dynamically evolves in response to external shocks, and that this structural change itself serves as a significant source of predictive information. Predictable patterns emerge as the market structure reorganizes in response to specific policy changes or macroeconomic events, and our model learns these patterns to forecast future returns. This suggests that to predict the price of an individual asset, it is not only the asset's historical data that is important, but it is also essential to analyze the dynamics of the entire market system. Consequently, this research presents new clues to predictability previously overlooked through the dynamic analysis of market structure, indicating that the analysis of exogenous variables can play a pivotal role in individual asset prediction.

The remainder of the paper is structured as follows. The next section reviews prior studies that apply sentiment analysis to stock market prediction and highlights recent developments in transformer-based hybrid modeling. Section 3 details the stock data employed in the analysis and outlines the procedure for collecting the news data. Section 4 describes the sentiment analysis framework and introduces the proposed models, including the baseline TVP-VAR and Transformer-based approaches. Section 5 reports the empirical forecasting outcomes and compares them with benchmark models. Finally, Section 6 concludes with a summary of the main findings and their implications.

2. Literature review

Traditional forecasting in finance has relied on econometric models such as Autoregressive Integrated Moving Average (ARIMA), Generalized AutoRegressive Conditional Heteroskedasticity (GARCH), and VAR to capture price dynamics and volatility. While effective under stable conditions, these models focus solely on numerical data and often fail to account for investor sentiment or regime shifts, limiting their predictive power. Our study builds on this limitation by incorporating additional non-price signals into the forecasting process.

One of the most important advances in this regard is the use of sentiment extracted from textual sources such as financial news, social media, and analyst reports. Early studies relied on sentiment dictionaries and Term Frequency–Inverse Document Frequency (TF-IDF) features; for example, [30] showed that aggregated sentiment from StockTwits microblogs improved directional accuracy compared to price-only models. Later, deep learning models moved sentiment analysis beyond simple lexical counts: [31] incorporated a CNN-derived sentiment index into an LSTM model, and [32] combined technical indicators with dictionary-based sentiment features using a two-layer LSTM. These studies consistently demonstrated that textual sentiment adds value to forecasting. However, their sentiment representation was often shallow, leaving much of the linguistic nuance unused. This gap motivated our focus on Transformer-based embeddings, which capture richer contextual information from financial language.

The development of domain-specific language models, particularly FinBERT, has substantially advanced financial sentiment analysis. For instance, [33] used FinBERT to classify sentiment in news headlines and showed that incorporating these features into an LSTM significantly improved prediction accuracy over price-only baselines. Similarly, [34] combined forum-based sentiment with technical indicators in a hybrid LSTM, again finding that models with sentiment consistently outperformed those without. These works establish two points: (i) financial sentiment carries unique predictive content, and (ii) domain-tuned Transformers like FinBERT are especially effective at extracting it. Yet, in most cases, sentiment is treated as a supplementary feature, without being embedded into a broader view of market structure. Our study extends this literature by not only using FinBERT embeddings but also situating them alongside dynamic measures of inter-market connectedness.

At the same time, research on hybrid deep learning architectures has shown the promise of Transformers in financial forecasting. The self-attention mechanism enables the capture of long-range dependencies and contextual interactions that LSTMs and CNNs often miss. For example, [35] adapted a Multi-Transformer model for volatility forecasting and found it superior to both traditional econometric models and LSTM-based hybrids. Similar hybrid architectures have been validated in various domains, from engineering systems [36] to intelligent transportation [37], underscoring the versatility of Transformer-based approaches. In finance, [38] showed that combining Bidirectional LSTM (BiLSTM), a modified Transformer with temporal convolution, and a Temporal Fusion Transformer outperformed baseline models in Shanghai and Shenzhen markets. These results collectively suggest that Transformer-based hybrids can effectively handle complex temporal data. However, most Transformer hybrids in finance rely solely on price data or technical indicators, without leveraging external sentiment or structural information. This motivates our contribution: A Transformer-based framework that integrates heterogeneous data streams.

Beyond sentiment, another important strand of the literature emphasizes modeling time-varying interdependencies across financial markets. The TVP-VAR framework has been widely employed to capture dynamic spillovers, particularly during periods of market stress or policy shifts. For example, [20] analyze volatility transmission in European sovereign bond markets using TVP-VAR, while [21] examine cross-market linkages under global uncertainty. More recent studies, such as [22] and [23], further demonstrate that TVP-VAR effectively tracks the evolving interactions among asset classes and across international markets.

Alongside TVP-VAR–based approaches, alternative frameworks have been proposed to model cross-market dependence. In particular, multivariate volatility models such as Dynamic Conditional Correlation GARCH (DCC-GARCH) are widely used to capture time-varying correlations across assets [39, 40]. These models offer a parsimonious and statistically tractable representation of co-movement dynamics and are well suited for volatility analysis. However, because DCC-GARCH primarily focuses on second-moment correlations, it does not explicitly characterize directional spillovers or network-based shock transmission, limiting its ability to capture asymmetric propagation mechanisms across markets [41].

Another influential line of research relies on connectedness measures derived from vector autoregressive frameworks to assess inter-market linkages and systemic risk [41, 42]. These measures provide intuitive interpretations of spillover intensity and are effective for monitoring market interconnectedness. Nevertheless, when implemented using static or rolling-window specifications, they may suffer from sensitivity to window selection and are often better suited for descriptive analysis than for modeling smoothly evolving market structures [43].

Table 2 compares alternative cross-market spillover modeling approaches and motivates the choice of the TVP-VAR framework. While DCC-GARCH and VAR-based connectedness measures provide useful insights into market dependence, TVP-VAR offers greater flexibility by allowing spillover dynamics to evolve continuously over time. This property makes TVP-VAR particularly well suited for capturing structural changes under heightened uncertainty and for integration into predictive forecasting models. Accordingly, this study embeds TVP-VAR–derived spillover measures as structured input features within a hybrid Transformer architecture, enabling joint modeling of dynamic market structure, sentiment, and price information.

Table 2. Comparison of cross-market spillover modeling approaches and justification of TVP-VAR.

Approach	Representative Studies	Main Strengths	Limitations for Forecasting
DCC-GARCH	[39, 40]	Captures time-varying correlations and volatility co-movements in a parsimonious multivariate framework	Focuses on second-moment dependence; lacks directional spillover interpretation and network-based transmission mechanisms [41]
VAR-based connectedness measures	[41–43]	Provides intuitive measures of spillover intensity and systemic risk; useful for monitoring inter-market linkages	Often relies on static or rolling-window estimation; sensitive to window choice and primarily descriptive in nature
TVP-VAR spillover models	[20–23]	Allows coefficients and spillovers to evolve smoothly over time; captures directional and network-based interdependencies, especially during stress periods	Typically used for spillover quantification and risk monitoring rather than embedded as predictive features
This study	–	Leverages TVP-VAR to extract dynamic structural information reflecting evolving market interdependencies	Embeds TVP-VAR spillover estimates as structured input features within a hybrid Transformer architecture, enabling joint modeling with sentiment and price information

While existing studies have advanced financial forecasting through sentiment analysis, econometric modeling, and deep learning, several methodological limitations remain. Sentiment-based approaches, including FinBERT-style language models, may suffer from domain shift across markets and time periods and can be sensitive to noisy or ambiguous textual information, particularly during turbulent market conditions. Likewise, TVP-VAR and related models are primarily used for descriptive analysis of dynamic spillovers and entail substantial computational and estimation costs, with scalability challenges as the dimensionality of the system increases. In addition, Transformer-based forecasting models, while powerful in capturing long-range dependencies, often rely on large parameter spaces and high-dimensional feature representations, raising concerns regarding overfitting, interpretability, and robustness when heterogeneous inputs are weakly structured. Moreover, many existing frameworks lack a clear formalization of how heterogeneous data sources are integrated and bounded within the model. These limitations motivate the need for a more critically grounded and systematically integrated approach that carefully balances model expressiveness with interpretability, robustness, and computational feasibility.

Recent advancements in financial forecasting have increasingly shifted towards multimodal deep learning frameworks that integrate heterogeneous data sources beyond historical prices. For instance, [44] demonstrated that fusing news sentiment with external shock variables, such as COVID-19 statistics, significantly enhances stock trend prediction. Similarly, [45] introduced an Investor Concern Sentiment Index by combining news text with search volume data to capture investor attention, while [46] utilized spatial feature fusion to integrate product sentiment with historical order data.Alongside text-based approaches, visual data has also emerged as a valuable predictive resource; for instance, [47] proposed a hybrid LSTM-CNN model incorporating candlestick charts to capture technical analysis features.

Despite these progressions, existing multimodal studies predominantly focus on micro-level factors or static visual patterns, often overlooking the dynamic evolution of inter-market dependencies. Unlike [47], which visualizes single-asset technical indicators, our study utilizes TVP-VAR to generate spillover heatmaps representing the macroscopic connectivity between markets. By integrating these structural visual features with micro-level sentiment via a hybrid Transformer, we establish a novel framework that explicitly captures the dynamic reorganization of market structures, a dimension largely unexplored in prior multimodal research.

Furthermore, despite the technical progress achieved in multimodal forecasting methodologies, several important issues remain concerning the structural design and implementation of the proposed framework.

First, regarding the level of detail in the text, full-text articles can provide richer contextual information; however, in practice they often contain unnecessary narration, redundancy, and off-topic sentences, thereby increasing preprocessing costs [48, 49]. By contrast, news headlines convey core events in a highly compressed form and are thus reported to be relatively less noisy while retaining immediate signals relevant to market reactions. Accordingly, headline-based text mining has been utilized in short-horizon financial fluctuation forecasting [33, 50].

Second, cross-market heterogeneity constitutes a critical challenge because market responses to news are not uniform across differing regulatory environments [51,52]. Therefore, accurately capturing the dynamics of each market generally requires country-specific data collection strategies rather than relying solely on a unified global news feed [53, 54].

Third, sentiment measurement is structurally influenced by media-specific reporting biases, particularly a negativity bias in financial news. Prior studies document that negative events tend to be reported more frequently and more prominently than positive ones, and that such negative tone is significantly associated with market trading activity [55, 56]. Consequently, sentiment indicators based only on simple polarity frequency (i.e., lexicon-based counting) may be overly skewed toward pessimism [57, 58]. To address this asymmetry, we avoid a purely lexicon-based approach and instead adopt a context-aware model such as FinBERT, which can capture nuanced semantic meanings in financial language, thereby mitigating the risk of distortions in the sentiment distribution [59, 60].

In addition, we provide a literature integration table (Table 3) that systematically summarizes existing studies, identifies their key limitations and research gaps, and clearly explains how the proposed framework addresses these gaps.

Table 3. Integration of prior literature and positioning of the present study.

Research Stream	Representative Studies	Key Limitations / Gaps	How This Study Differs / Contributes
Sentiment-based forecasting	[30–32]	Early sentiment representations rely on dictionary-based or shallow embedding methods, which limit the ability to capture contextual and nuanced financial language.	Utilizes Transformer-based sentiment embeddings to extract richer contextual information from financial text.
FinBERT-based sentiment models	[33, 34]	Although FinBERT improves sentiment extraction, sentiment features are typically treated as auxiliary inputs and are not integrated with macro-level market structure.	Integrates FinBERT-based micro-level sentiment with dynamic market structure in a unified forecasting framework.
Hybrid deep learning and Transformer models	[35–38]	Most Transformer-based hybrids remain price-centric or rely on technical indicators, with limited use of external sentiment or structural information.	Employs a Transformer architecture with self-attention to jointly model interactions among price, sentiment, and structural features.
Dynamic market structure and spillovers (TVP-VAR)	[20–23]	TVP-VAR outputs are primarily used for descriptive spillover analysis and systemic risk monitoring rather than as predictive inputs.	Transforms TVP-VAR spillover estimates into heatmaps and embeds them as predictive features within a hybrid Transformer model.
Multimodal financial forecasting	[44–47]	Existing multimodal approaches focus mainly on micro-level sentiment or static visual features and largely overlook time-varying inter-market connectivity.	Introduces a multimodal framework that fuses sentiment, price, and TVP-VAR–based structural connectivity to capture dynamic market reorganization.
This study	–	–	Proposes a hybrid Transformer framework that systematically integrates FinBERT-based sentiment, TVP-VAR–based dynamic market structure, and historical prices for enhanced forecasting accuracy.

Consequently, in light of the limitations identified in the existing literature reviewed in this section, the present study seeks to address the following research gaps. First, much of the existing literature relies on single-source inputs or loosely coupled multimodal approaches, typically incorporating investor sentiment or market structure separately, which limits their ability to jointly capture behavioral and systemic drivers of asset prices. Second, while time-varying parameter models such as TVP-VAR are widely used to analyze dynamic spillovers and systemic risk, their outputs are primarily treated as descriptive diagnostics rather than as structured predictive features embedded within forecasting models. Third, although Transformer-based architectures have demonstrated strong performance in financial time-series prediction, most applications remain heavily price-centric and lack a rigorous framework for integrating heterogeneous data sources with clearly defined transformation and interaction assumptions. These gaps motivate the present study, which proposes a unified hybrid Transformer framework that systematically integrates FinBERT-based sentiment measures and TVP-VAR-based dynamic market structure with historical price information to enhance predictive accuracy and better reflect the adaptive and interconnected nature of financial markets.

3. Data description

3.1. Stock index

This study utilizes daily closing prices from four major benchmark equity indices: the S&P 500 (Standard & Poor's 500 Index, United States), the FTSE 100 (Financial Times Stock Exchange 100 Index, United Kingdom), the CSI 300 (China Securities Index 300, China), and the Nikkei 225 (Nikkei 225 Stock Average, Japan). The data were retrieved from the `Investing.com` database. The sample period spans from 1 January 2020 to 31 December 2024, a horizon selected to capture post-COVID-19 structural changes in global equity markets and to mitigate regime-related biases associated with pre-pandemic observations. All entries with missing values were removed, after which logarithmic returns were computed for each index. Subsequently, basic statistical analyses were conducted to characterize the time-series properties. These preprocessing procedures provide the empirical foundation for the analyses that follow.

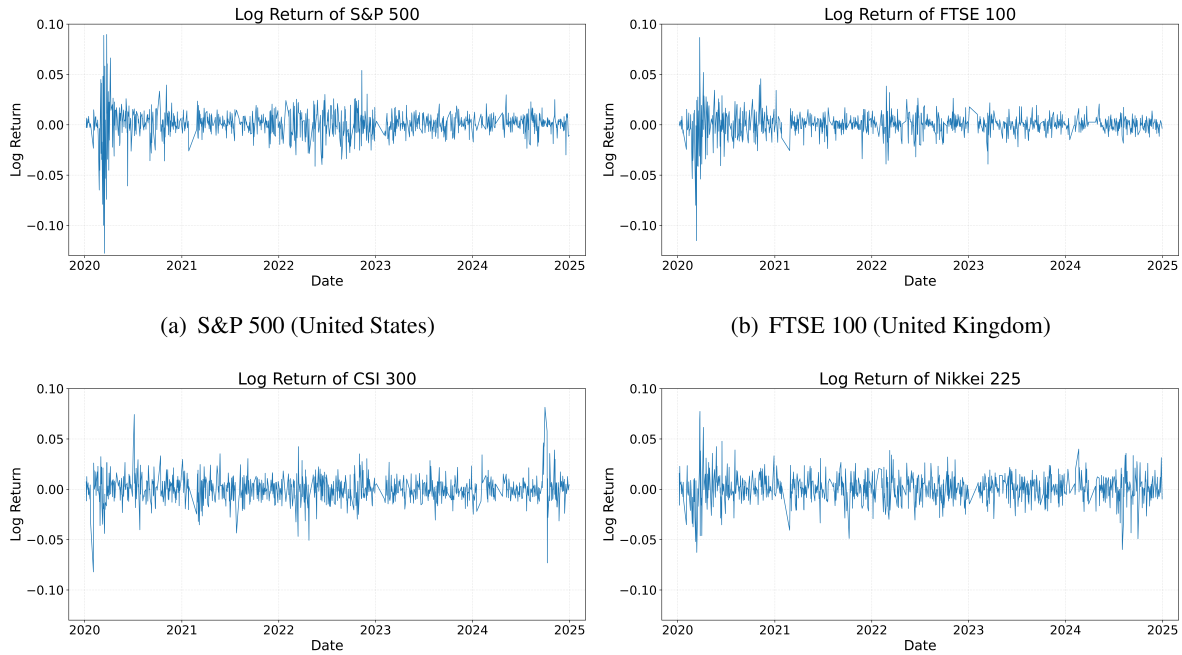
Table 4 presents the basic descriptive statistics and stationarity test outcomes for the daily log returns of four national equity indices. Overall, the mean returns are near zero, with the S&P 500 exhibiting the highest average at 0.0006 and the FTSE 100 the lowest at 0.0001, suggesting that most indices display long-term mean-reverting characteristics.

Table 4. Basic summary statistics and test results for the country-level index returns. The Jarque–Bera statistic (J.-B.) tests the null hypothesis of normality for the sample returns; ADF and PP are the Augmented Dickey–Fuller and Phillips–Perron unit-root tests, respectively. ‡ denotes rejection of the null hypothesis at the 1% significance level.

Index	Obs	Mean	Max.	Min.	Std. dev.	Skew	Kurt.	J.-B.	ADF	PP
S&P 500	1071	0.0006	0.0897	-0.1277	0.0143	-0.9545	16.8430	8714.13‡	-9.81‡	-38.79‡
FTSE 100	1071	0.0001	0.0867	-0.1151	0.0117	-1.0878	18.3552	10732.96‡	-8.09‡	-34.26‡
CSI 300	1071	0.0001	0.0814	-0.0821	0.0131	0.0696	8.4607	1331.55‡	-14.59‡	-30.97‡
Nikkei 225	1071	0.0005	0.0773	-0.0627	0.0138	-0.0138	6.0263	408.74‡	-21.24‡	-32.61‡

An analysis of extreme values reveals that the S&P 500 exhibits the widest range, with a maximum return of 0.0897 and a minimum of -0.1277. In contrast, the Nikkei 225 records the narrowest span, ranging from a maximum of 0.0773 to a minimum of -0.0627.

Figure 1 provides a time series visualization of the daily log returns for each index from January 2020 to December 2024. Across all indices, a marked surge in volatility is observable during the initial phase of the COVID-19 pandemic, followed by a return to relatively stable levels. The S&P 500, in particular, exhibits visually pronounced spikes during early 2020, which align with their high kurtosis and deviations from normality. Meanwhile, Nikkei 225 and CSI 300 display more constrained fluctuation patterns. These temporal dynamics underscore the presence of heteroskedasticity and asymmetry, complementing the statistical diagnostics.



(a) S&P 500 (United States)

(b) FTSE 100 (United Kingdom)

(c) CSI 300 (China)

(d) Nikkei 225 (Japan)

Figure 1. Time-series plots of daily log returns for four major stock indices from 1 January 2020 to 31 December 2024. Each panel illustrates the temporal evolution of log returns for: (a) S&P 500 (United States), (b) FTSE 100 (United Kingdom), (c) CSI 300 (China), and (d) Nikkei 225 (Japan).

In terms of volatility, the standard deviations of the S&P 500 (0.0143), CSI 300 (0.0131), and Nikkei 225 (0.0138) are broadly similar, whereas the FTSE 100 (0.0117) reveals comparatively lower volatility.

The skewness and kurtosis metrics further indicate that most return distributions are asymmetric and fat-tailed. For instance, the FTSE 100 demonstrates a pronounced left skew (–1.0878) and extreme kurtosis (18.3552), implying frequent occurrences of tail events. These statistical properties are corroborated by the J.-B. test, which decisively rejects the null of normality for all indices at the 1% significance level.

Additionally, the results of the ADF and PP unit root tests show that the null hypothesis of a unit root is rejected at the 1% significance level for all series, indicating that the log returns of all indices satisfy the stationarity condition. This serves as a fundamental prerequisite for the validity of subsequent econometric analyses.

In summary, although the daily log returns of the four indices converge toward zero on average, their return distributions deviate notably from normality, particularly in terms of tail risk and asymmetry. Moreover, the confirmation of stationarity across all series affirms the statistical soundness of the empirical framework that follows.

3.2. News data

The news data used in this study were collected from publicly available article listings on Reuters (`https://www.reuters.com`). For dates prior to 2023, we used archive pages with the path pattern `/archive/{YYYY}-{MM}/{DD}/{page}`, whereas for dates from 2023 onward we used sitemap pages with the path pattern `/sitemap/{YYYY}-{MM}/{DD}/{page}`. The target indices were S&P 500, FTSE 100, CSI 300, and Nikkei 225. Only article titles were used to conduct sentiment classification and daily aggregation.

The news collection pipeline was implemented using `Selenium` with `undetected-chromedriver`. We collected articles only on trading days, excluding exchange holidays for each index. From Reuters archive listings, we filtered candidate documents using the keyword set provided in Appendix 6.3. For each target date, headlines were extracted via CSS selectors. To mitigate server load and avoid rate limiting, we introduced inter-request delays, incremental scrolling, and bounded retries. Extracted headlines underwent light normalization and de-duplication, after which they were stored by index according to a predefined keyword index mapping.

We collected headline data for the set of calendar dates common across all indices from 2020-01-06 through 2024-12-27. To ensure reliable daily aggregation, we applied a low-volume filter: If no more than 5 headlines were available on a given date, we deemed the daily state indeterminate and excluded that date from all indices. This filter was applied uniformly to maintain comparability across indices.

Table 5 reports the volume of news articles by index and the corresponding sentiment analysis of news titles obtained using FinBERT. Among the indices, the largest number of news titles was collected for the S&P 500 (81314 titles), whereas the smallest was for the Nikkei 225 (9245 titles). Across all indices, the sentiment distribution generally follows the order positive, followed by neutral, and then negative.

Table 5. Distribution of sentiment polarity in the news dataset: Descriptive statistics of Reuters headlines classified via FinBERT.

Index	Total	Positive	Neutral	Negative	Daily Average
S&P 500	81314	32077	30700	18537	96.5
FTSE 100	25082	9931	9267	5884	29.8
CSI 300	20242	7664	7059	5519	24
Nikkei 225	9245	3479	3313	2453	11

3.3. Preliminary diagnostics and lag selection

To validate the structural appropriateness of the proposed framework and capture the dynamic properties of the time-series data, we conducted preliminary diagnostic tests on the daily log returns. Specifically, the Ljung–Box test was employed to detect the presence of serial autocorrelation, while the AutoRegressive Conditional Heteroskedasticity Lagrange Multiplier (ARCH-LM) test was utilized to examine conditional heteroskedasticity.

According to Table 6, the S&P 500 and FTSE 100 indices exhibited strong autocorrelation at the 1% significance level, whereas the CSI 300 and Nikkei 225 displayed relatively weak linear dependence. However, the results of the ARCH-LM test, employed to verify volatility clustering, indicate that the null hypothesis was decisively rejected for all four indices at the $p < 0.001$ level. This finding implies the distinct presence of conditional heteroskedasticity in the data, thereby statistically validating the application of the TVP-VAR framework—which captures time-varying volatility—over conventional models relying on fixed parameters.

Table 6. Preliminary diagnostics for autocorrelation and conditional heteroskedasticity. The table reports the test statistics and corresponding p-values for the Ljung–Box test and the ARCH-LM test on daily log returns.

	Autocorrelation (Ljung–Box)		Heteroskedasticity (ARCH-LM)	
Index	Statistic $Q(10)$	p-value	Statistic $LM(10)$	p-value
S&P 500	107.90	0.000	422.17	0.000
FTSE 100	27.85	0.002	270.62	0.000
CSI 300	16.07	0.098	132.32	0.000
Nikkei 225	17.36	0.067	182.08	0.000

To determine the optimal lag length for the TVP-VAR model, we examined standard information criteria including the Akaike Information Criterion (AIC), Bayesian Information Criterion (BIC), and Hannan-Quinn Information Criterion (HQ). As shown in Table 7, the AIC suggests a lag order of 2 as optimal (with the lowest value of -35.504). However, both the BIC and HQ, which impose stricter penalties on model complexity, favor a lag order of 1 (with values of -35.359 and -35.432, respectively). Given the risk of overfitting inherent in time-varying parameter models, we prioritized the principle of parsimony supported by the BIC and HQ criteria. Consequently, we set the lag order to $p = 1$ for the subsequent analysis.

Table 7. Lag-order selection for the VAR system using information criteria. Bold values indicate the optimal lag suggested by each criterion.

Lag p	AIC	BIC	HQ
1	-35.478	-35.359	-35.432
2	-35.504	-35.289	-35.421
3	-35.477	-35.166	-35.358
4	-35.452	-35.045	-35.296

4. The proposed model

4.1. Sentiment analysis

In this study, we incorporate sentiment information extracted from financial news as an auxiliary feature to enhance the prediction of log-returns for country-level stock indices. To this end, we adopt FinBERT, a domain-specific, transformer-based pre-trained language model tailored for financial text analytics. Traditional return prediction models have predominantly relied on structured data such as macroeconomic indicators and technical factors [61,62], while recent studies underscore the predictive value of investor sentiment derived from unstructured textual data [63,64]. Particularly, contextual cues embedded in global financial news reflecting country-specific macroeconomic conditions, monetary policies, and geopolitical risks can significantly influence investor expectations and sentiment, thereby shaping short to medium term volatility in national equity markets [65, 66].

FinBERT is a variant of the BERT architecture, re-trained and fine-tuned specifically for the financial domain. Its training pipeline consists of three stages, as depicted in Figure 2. The first stage involves generic-domain pretraining using large scale textual corpora, where FinBERT learns deep contextual representations via Masked Language Modeling (Masked LM, or MLM) and Next Sentence Prediction (NSP). The pretraining loss is formally defined as:

$$\mathcal{L}_{\text{pretrain}} = \mathcal{L}_{\text{MLM}} + \mathcal{L}_{\text{NSP}} \,. \tag{4.1}$$

In Eq (4.1), $\mathcal{L}_{\text{MLM}}$ denotes the cross-entropy loss for predicting masked tokens, and $\mathcal{L}_{\text{NSP}}$ captures binary classification loss for sentence pair coherence.

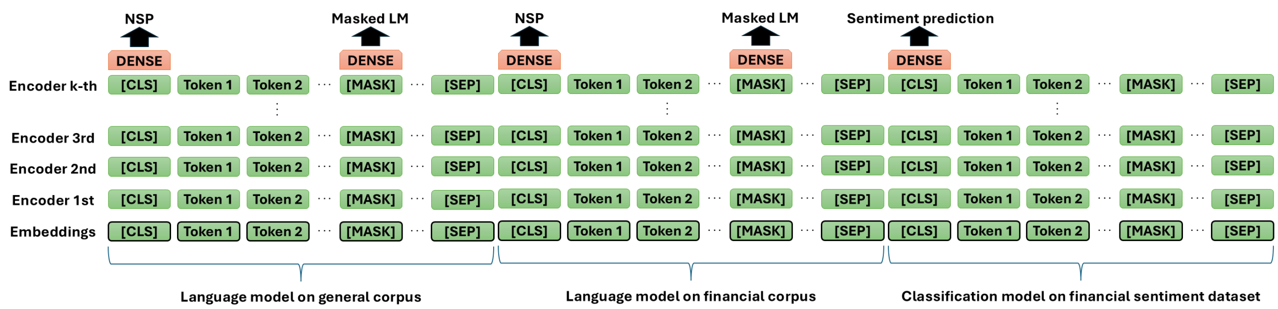


Figure 2. The basic architecture of the FinBERT model.

In the second stage, domain-adaptive pretraining is performed on financial corpora, enabling the model to capture domain-specific semantics, such as terms like "quantitative easing", "credit downgrade", and "rate hike", which are frequent in economic discourse but rare in general purpose datasets.

The final stage involves task-specific fine-tuning for sentiment classification. Given an input sentence, FinBERT extracts the contextual embedding [CLS] corresponding to the [CLS] token and passes it through a softmax classification layer:

$$\hat{y} = \mathrm{softmax}(Wh_{[\mathrm{CLS}]} + b) \;, \tag{4.2}$$

$$\mathcal{L}_{\mathrm{sent}} = -\sum_{c=1}^{C} y_c \log(\hat{y}_c) \;. \tag{4.3}$$

In Eq (4.2), the model computes the predicted probability distribution, $\hat{y}$, over the sentiment classes. This is achieved by feeding the final hidden state vector $h_{[\mathrm{CLS}]} \in \mathbb{R}^d$, which corresponds to the special [CLS] token and serves as an aggregate representation of the input sequence, into a classification layer. This layer consists of a trainable weight matrix $W \in \mathbb{R}^{C\times d}$ and a bias vector $b \in \mathbb{R}^C$. The model is then optimized by minimizing the categorical cross-entropy loss $\mathcal{L}_{\mathrm{sent}}$, as defined in Eq (4.3). In this loss function, $\hat{y}_c$ is the predicted probability for a given class c, and y_c is the one-hot encoded ground-truth label (i.e., $y_c = 1$ for the true class and 0 otherwise). Here, $C = 3$ represents the number of classes (positive, neutral, negative), and d is the dimension of the hidden state.

Through its context-aware representation capability, rich financial vocabulary coverage, and document-level sentiment integration, FinBERT allows the extraction of quantitative sentiment indicators from unstructured news text. These sentiment features are expected to play a complementary role in enhancing the explanatory and predictive performance of traditional price-based return models.

4.2. TVP-VAR model

TVP-VAR is a structural time-series framework that extends the conventional VAR model by permitting its coefficients to evolve dynamically over time.

This flexibility enables the model to capture dynamic interactions that static specifications struggle to explain in financial markets, where structural breaks and non-stationarity are commonplace [67,68].

The standard VAR(p) model is described by Eq (4.4):

$$y_t = A_1 y_{t-1} + A_2 y_{t-2} + \cdots + A_p y_{t-p} + \varepsilon_t \;, \tag{4.4}$$

where $y_t \in \mathbb{R}^n$ denotes an n-dimensional vector of time-series variables, $A_i \in \mathbb{R}^{n\times n}$ are fixed coefficient matrices, and $\varepsilon_t \sim \mathcal{N}(0, \Sigma)$ is white noise.

In the TVP-VAR specification, the coefficient matrices are allowed to vary with time, leading to the reformulation in Eq (4.5):

$$y_t = A_{1,t} y_{t-1} + A_{2,t} y_{t-2} + \cdots + A_{p,t} y_{t-p} + \varepsilon_t \;, \tag{4.5}$$

and the temporal evolution of each coefficient is modeled as the random-walk process in Eq (4.6):

$$\mathrm{vec}(A_{i,t}) = \mathrm{vec}(A_{i,t-1}) + \eta_{i,t} \;, \qquad \eta_{i,t} \sim \mathcal{N}(0, Q_i) \;. \tag{4.6}$$

This setup naturally accommodates the time-inconsistency generated by policy shifts or market instability [3, 69] and is well suited for tracing shock transmission and dynamic connectedness across financial markets [70, 71].

In this study, we apply TVP-VAR to country-level stock index log returns to compute time-varying connectivity strengths, visualize them as spillover heatmaps, embed the resulting images using a CNN, and integrate these embeddings into the Transformer input. This approach implements a multi-modal learning architecture that combines structured time-series data with visualized network features.

4.3. *Transformer model*

The Transformer model is a non-sequential, encoder-decoder architecture that eliminates sequential recursive operations, thereby enabling the entire input sequence to be processed in parallel. Owing to this design, it has been widely adopted not only in natural-language processing but also in domains such as financial time-series forecasting [72, 73]. The baseline configuration employed in the present study is illustrated in Figure 3.

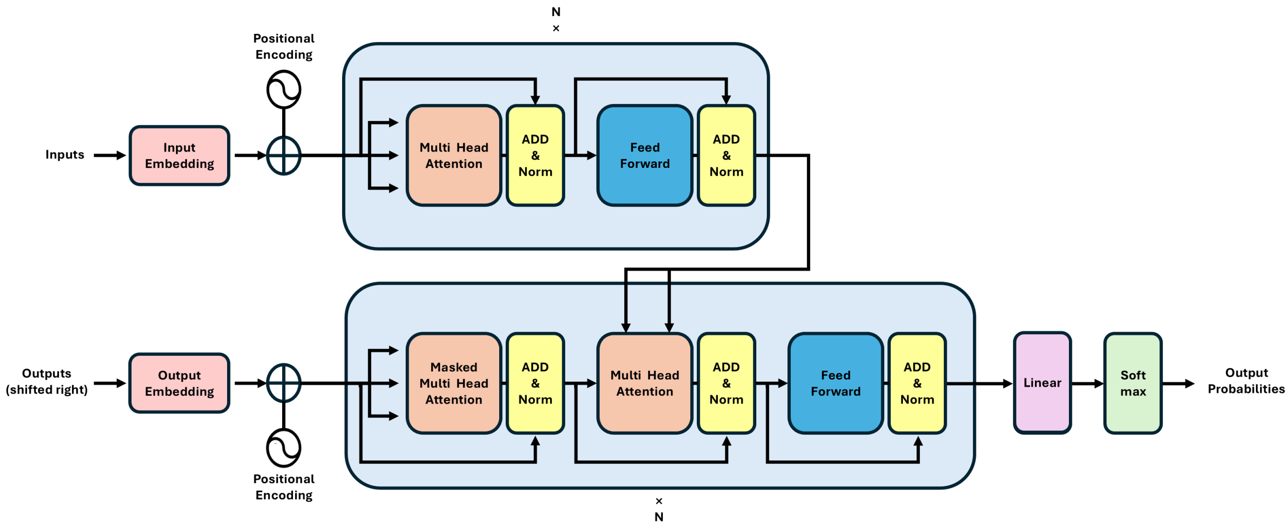


Figure 3. The basic architecture of the Transformer model.

The input sequence $x = (x_1, x_2, \ldots, x_T)$ is first mapped to a d-dimensional embedding, yielding $E \in \mathbb{R}^{T\times d}$. A deterministic positional-encoding matrix $P \in \mathbb{R}^{T\times d}$ is then added to form the initial representation $Z^{(0)} = E + P$. This matrix $Z^{(0)}$ serves as the input to the first encoder block. In general, the output of the $(l-1)$-th block, denoted as $Z^{(l-1)}$, becomes the input for the l-th block.

Each of the N encoder blocks consists of a Multi-Head Self-Attention (MHSA) module followed by a position-wise feed-forward network. In Eq (4.7) scaled dot-product attention is defined as :

$$\text{Attention}(Q, K, V) = \text{softmax}\left(\frac{QK^{\top}}{\sqrt{d_k}}\right)V\,. \tag{4.7}$$

In Eq (4.7), Q, K, and V denote the Query, Key, and Value matrices, respectively. The term d_k represents the dimension of the Key and Query vectors. The dot product is scaled by a factor of $\frac{1}{\sqrt{d_k}}$ to prevent it from growing too large, which could push the softmax function into regions with extremely small gradients, thereby stabilizing the training process. For a given encoder block receiving the input representation $Z \in \mathbb{R}^{T\times d}$, these are computed as linear projections: $Q = ZW^Q$, $K = ZW^K$, and $V = ZW^V$. Here, W^Q, W^K, and $W^V \in \mathbb{R}^{d\times d_k}$ are the trainable weight matrices. The superscripts Q, K, and V are not mathematical exponents but rather notational identifiers, indicating that these matrices are used to project the input into the Query, Key, and Value spaces, respectively.

In Eq (4.8) the MHSA, parallelized over h heads, is formulated as:

$$\mathrm{MultiHead}(Q, K, V) = \mathrm{Concat}(\mathrm{head}_1, \ldots, \mathrm{head}_h)W^O \,, \tag{4.8}$$

$$\mathrm{head}_i = \mathrm{Attention}(QW_i^Q, KW_i^K, VW_i^V) \,. \tag{4.9}$$

In this formulation, the projection matrices for each head are distinct: $W_i^Q \in \mathbb{R}^{d\times d_k}$, $W_i^K \in \mathbb{R}^{d\times d_k}$, and $W_i^V \in \mathbb{R}^{d\times d_v}$, where it is common to set $d_k = d_v = d/h$. In Eq (4.8), $W^O \in \mathbb{R}^{hd_v\times d}$ denotes the output projection matrix. Each head, defined in Eq (4.9), captures dependencies in distinct sub-spaces; the concatenated heads are subsequently projected to form an integrated representation. Every sub-layer is equipped with residual connections and layer normalization to enhance training stability and mitigate vanishing gradients.

The decoder mirrors the encoder architecture, except that its first attention sub-layer employs masking to prevent information leakage from future positions. The subsequent encoder-decoder attention module conditions the predictions on the encoder output.

The final decoder state is transformed into the output distribution according to

$$\hat{y}_t = \mathrm{softmax}(Wh_t + b) \,. \tag{4.10}$$

In Eq (4.10), $h_t \in \mathbb{R}^d$ denotes the final output representation of the decoder stack for the t-th position, $W \in \mathbb{R}^{V\times d}$ is the output projection matrix, and V is the vocabulary (or class) size.

By obviating sequential dependencies, the Transformer achieves substantial training-speed gains through parallelization [74, 75] and effectively alleviates the long-term dependency vanishing-gradient problem inherent to RNN-based models [76, 77].

4.4. Theoretical rationale for model design

In this study, we conceptualize financial markets not as static equilibrium systems but as time-varying dynamic complex systems in which information-transmission pathways are continuously reconfigured by exogenous shocks and investor sentiment. From the perspective of the Adaptive Markets Hypothesis (AMH), market efficiency and risk-transmission channels are not fixed constants; rather, they evolve in response to changing environments. Accordingly, the forecasting task is framed not as a simple extrapolation of a univariate time series but as the approximation of a mapping that generates next-period returns given the heterogeneous information set observable up to time t. In our setting, the information set encompasses (i) trend and volatility signals extracted from past prices, (ii) sentiment and uncertainty signals reflected in news, and (iii) changes in cross-country connectivity structure that characterize systemic risk transmission.

To present this framework in a reproducible and verifiable form, we decompose model inputs into three channels—Price, Sentiment, and Structure—and explicitly fix the data-transformation rules for each channel. First, the Sentiment channel converts daily news headlines into time-series variables by applying FinBERT to obtain sentiment distributions and confidence scores, followed by date-level aggregation. Importantly, the aggregation procedure is implemented as a fixed rule applied identically across all countries to ensure reproducibility. Second, the Structure channel assumes that interactions in a multivariate return system evolve over time; we therefore estimate time-specific connectivity strengths via a TVP-VAR model and transform them into a Generalized Forecast Error Variance Decomposition (GFEVD)-based spillover matrix. We interpret this spillover matrix not as a mere summary statistic but as a weighted network representation of the time-dependent systemic risk transmission structure. Third, the Price channel consists of a fixed-length historical log-return sequence, and all training and inference are time-aligned so that only information available prior to time t is used.

In particular, the CNN-on-heatmaps component is defined as a representation-learning step designed to robustly extract structural patterns embedded in spillover matrices, rather than as a visualization trick. Concretely, each spillover matrix is (1) normalized using an identical rule, (2) converted into a standardized input format with fixed resolution, and (3) compressed into a low-dimensional embedding via a CNN. This standardization is not simply making an image; it is a transformation intended to preserve, in a consistent manner, structural characteristics implied by the relative arrangement of matrix elements—such as country-centered concentration of transmission, asymmetric spillovers, and clustered co-movement—within the model's input space. As a result, the Structure channel provides a numerical embedding of the time-varying network state that can be effectively processed by the Transformer.

Moreover, features from the three channels are merged on a shared timeline to form an integrated input sequence. The Transformer's self-attention mechanism then learns, in a context-dependent manner, which factors (structural changes, sentiment signals, or price trends) are more informative for prediction at different periods. This design implements the AMH view that market participants adapt their decision rules as the environment changes. To maintain causal validity, all features—including spillover matrices and aggregated sentiment—are computed strictly using information available up to time t, and in the test period we apply only the transformation rules fixed in the training period. In sum, the proposed framework goes beyond an engineering-style concatenation of heterogeneous data by structurally translating the theoretical premise that micro-level sentiment and macro-level connectivity jointly shape prices into a principled pipeline of channel decomposition—fixed transformations—dynamic integration—causality constraints.

4.5. *Proposed model*

The architectural design of the proposed model is grounded in the theoretical premise that financial market dynamics are driven by the inter play of investor psychology, structural inter connectedness, and historical trends. We employ FinBERT to capture the 'Sentiment Channel', addressing the behavioral finance perspective that price data alone cannot fully reflect investor irrationality. Simultaneously, the TVP-VAR models the 'Structural Channel', based on the econometric theory that risk transmission pathways among markets evolve dynamically over time. In addition, the Transformer serves as the integration engine, selected for its ability to learn non-linear interactions and long-range dependencies among these heterogeneous features, thereby overcoming the limitations of traditional RNN-based models.

The model proposed in this study, as illustrated in Figure 4, is a hybrid, Transformer-based architecture designed to maximize predictive performance by systematically integrating multifaceted information. The model's operational principle consists of three main stages: Parallel feature extraction, multi-dimensional feature fusion and deep learning, and an ensemble-based final prediction.

Based on the hypothesis that asset price formation is driven by the dynamic interaction between micro-level investor psychology and macro-level systemic linkages, the proposed model employs a parallel feature extraction stage to capture these distinct dimensions [78, 79]. The model extracts core predictive information through three independent modules.

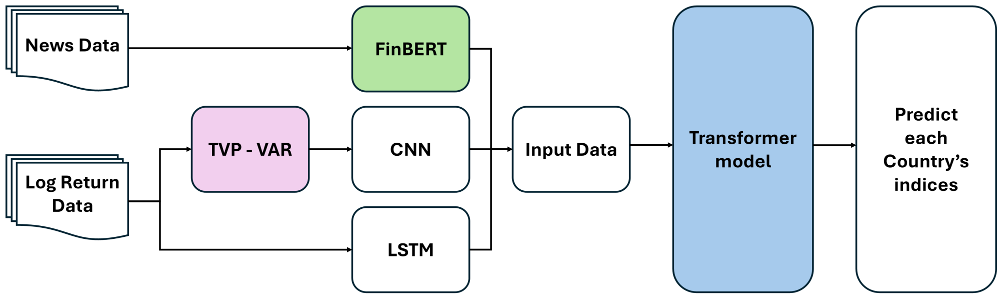


Figure 4. Architectural overview of the proposed hybrid Transformer framework integrating sentiment analysis and dynamic market structure.

First, the sentiment feature extraction module quantifies the market's psychological drivers from unstructured news data. We utilized the pre-trained FinBERT model ('ProsusAI/finbert'), which utilizes a BERT architecture specifically fine-tuned on the Financial PhraseBank dataset [80]. This benchmark corpus comprises 4,840 sentences extracted from financial news, manually annotated by 16 experts with financial backgrounds to ensure high inter-annotator agreement. Given that the model demonstrated superior performance with a classification accuracy of approximately 97% on this domain-specific dataset [81], we applied it directly for inference without additional fine-tuning. This approach allows us to leverage its validated capabilities while maintaining generalization across diverse global markets. FinBERT analyzes news headlines to extract sentiment scores that represent the positive, neutral, and negative sentiments of market participants. Specifically, as detailed in Algorithm 1, for each news item i, the model applies FinBERT (M_{FB}) to generate a Softmax probability distribution ($\mathbf{P}_i$). The class with the maximum probability is identified as the sentiment (Sentiment_i), while the probability value itself serves as the confidence score (Confidence_i). Subsequently, through a daily aggregation process at time t, the most frequent sentiment class is mapped to a daily sentiment score (S_t), and the average of the individual confidence scores (C_t) is computed to construct the final daily sentiment dataset ($\mathcal{D}_{\text{Sentiment}}$).

Second, the structural feature extraction module captures systemic changes across the entire market, moving beyond individual assets to account for risk contagion. The log return data of national stock indices are fed into a TVP-VAR model(lag over $p = 1$, horizon $H = 10$), which estimates how the interconnectivity and spillover effects between markets evolve over time. This dynamic connectivity information at each time point is converted into a two-dimensional image, such as a heatmap, and subsequently compressed into a high-dimensional feature vector using a CNN (comprising 3 convolutional layers with kernel size of 3 and 5). Following the procedure outlined in Algorithm 2, based on the log returns up to time t ($\mathbf{R}_{\log,1:t}$), the model estimates the time-varying coefficients ($\mathbf{A}_{l,t}$) and the covariance matrix ($\mathbf{\Sigma}_t$). To ensure robust parameter estimation under non-stationary conditions, we specifically employ a Kalman Filter with forgetting factors. These estimates are then transformed into spillover indices ($\mathbf{\Theta}_t$) via GFEVD. Spillover indices $\mathbf{\Theta}_t$ proxies the time-varying intensity of volatility spillovers; elevated connectedness reflects heightened market integration, implying diminished risk mitigation capacities. This dynamic connectivity information is visualized as a two-dimensional heatmap image (H_t) at each time step, which is subsequently compressed into a high-dimensional structural feature vector ($M_{k,t}$) via a CNN embedder (M_{CNN}) and a projector (M_{Proj}), thereby converting visual patterns of market structure into numerical features that represent the evolving systemic risk profile of the financial system. Crucially, this integration enables the Transformer to incorporate global structural context into its learning of historical price dependencies.

Algorithm 1 FinBERT-based news sensitivity extraction and aggregation ($\mathcal{A}_1$)

Require: News Headlines Data $\mathcal{D}_{\text{News}}$, Pre-trained FinBERT Model M_{FB} (ProsusAI)
Ensure: Daily Sentiment Score S_t and Average Confidence C_t

Initialize: Load M_{FB} tokenizer and model (no fine-tuning on downstream tasks)
for Each News Item i at Date t **do**
  $\mathbf{P}_i \leftarrow \text{Softmax}(M_{\text{FB}}(\text{Title}_i))$ ▷ Probability distribution over {Pos, Neg, Neu}
  $\text{Sentiment}_i \leftarrow \arg\max(\mathbf{P}_i)$
  $\text{Confidence}_i \leftarrow \max(\mathbf{P}_i)$
end for
for Each Unique Date t **do**
  $S_t \leftarrow \text{MapToValue}(\text{MostFrequent}(\{\text{Sentiment}_i \mid i \in \mathcal{R}_t\}))$
  $C_t \leftarrow \text{Average}(\{\text{Confidence}_i \mid i \in \mathcal{R}_t\})$
end for
return Daily Aggregated Sentiment $\mathcal{D}_{\text{Sentiment}} = \{(t, S_t, C_t)\}$

Algorithm 2 TVP-VAR and CNN based market structure embedding ($\mathcal{A}_2$)

Require: Log Returns $\mathbf{R}_{\log}$, Lag $p = 1$, Horizon $H = 10$, CNN M_{CNN}, Projector M_{Proj}
Ensure: Daily Market Structure Embeddings $M_{k,t} \in \mathbb{R}^{D_{\text{model}}}$

Hyperparams: TVP ($\delta = 0.99$, $\lambda = 0.99$), CNN ($L = 3$, $K = 3, 5$, $D_{\text{emb}} = 128$)
for Time Step t **do**
  $\mathbf{A}_{l,t}, \mathbf{\Sigma}_t \leftarrow \text{TVP-VAR}(\mathbf{R}_{\log,1:t}, p)$ ▷ Estimate time-varying params
  $\mathbf{\Theta}_t \leftarrow \text{GFEVD}(\{\mathbf{A}_{l,t}\}_{l=1}^{p}, \mathbf{\Sigma}_t, H)$ ▷ Calculate spillover index
  $H_t \leftarrow \text{GenerateHeatmap}(\mathbf{\Theta}_t)$ ▷ Visualize as 2D image
  for Asset k **do**
    $\mathbf{X}_{\text{Img},k,t} \leftarrow \text{ProcessHeatmap}(H_t, k)$ ▷ Masking for each asset
    $Z_{k,t} \leftarrow M_{\text{CNN}}(\mathbf{X}_{\text{Img},k,t})$ ▷ Extract visual features
  end for
  $\mathbf{Z}_t \leftarrow [Z_{1,t}, \ldots, Z_{N,t}]^{\intercal}$
  $\mathbf{M}_{\text{Tok},t} \leftarrow M_{\text{Proj}}(\mathbf{Z}_t)$ ▷ Project to D_{model} space
  $M_{k,t} \leftarrow \mathbf{M}_{\text{Tok},t}[k]$
end for
return Daily Market Structure Embeddings $\mathcal{D}_{\text{Market}} = \{(t, M_{k,t})\}$

The image-based encoding technique for spillover matrices adopted in this study possesses a theoretical grounding that transcends simple data transformation, effectively capturing complex inter dependencies within the market that vector-based approaches often overlook. The process of mapping time-series or matrix data into a two-dimensional image space preserves the correlations among variables as structural visual patterns, thereby contributing to the model's ability to learn non-linear relationships [82, 83]. Specifically, the convolutional operations of CNNs excel at analyzing local regions on a heatmap to identify specific clusters with high risk-transmission intensity or abrupt shifts in market structure as meaningful predictive features [84]. Consequently, by transforming dynamic market spillovers into spatial features, this approach provides a foundation for more precisely representing and analyzing economic connectedness [85].

Third, in the multi-dimensional feature fusion and deep learning stage, all parallelly extracted features-namely, the extracted sentiment scores, encoded market structure embeddings, and historical log returns are merged into a single multi-dimensional time-series sequence based on common calendar timestamps to ensure temporal alignment. As elaborated in Algorithm 3, the merged dataset ($\mathcal{D}_{\text{Features}}$) is processed via a sliding window technique to form batches ($\mathbf{X}$). To rigorously prevent data leakage, the target variable is shifted such that the model utilizes information strictly up to time t to predict returns at $t + 1$. Instead of simple concatenation, the proposed framework employs an LSTM layer to process the historical log return data, extracting intrinsic short-term temporal dependencies. These LSTM-encoded price features are then concatenated with the sentiment scores derived from FinBERT and the market structure embeddings encoded by the CNN. Finally, this fused feature vector is linearly projected to the model dimension and combined with positional embeddings to serve as the input for the Transformer encoder.

Within the Transformer encoder (configured with $N = 8$ layers, $h = 10$ attention heads, and hidden dimension $d_{\text{model}} = 320$), the MHSA mechanism captures global inter-dependencies among all elements within the input sequence. It dynamically learns which past features (sentiment, structure, or price) and which time steps are most crucial for future predictions, thereby enabling deep modeling of complex, long-range dependencies inherent in the time-series data.

Finally, the feature vector ($\mathbf{Z}_{\text{Enc, CLS}}$), now enriched with temporal context from the Transformer encoder, is passed to a Fully-Connected Layer to generate the final prediction ($\hat{\mathbf{Y}}$). During the training process, the model minimizes a composite loss function ($\mathcal{L}_{\text{Composite}}$) that simultaneously accounts for the error between predicted and actual values ($\mathbf{Y}$) as well as their correlation (Pearson Correlation). To improve robustness and control overfitting, we implemented comprehensive regularization strategies including Dropout ($p = 0.1$), Weight Decay ($\lambda = 1e - 4$) in the AdamW optimizer, and Early Stopping based on validation loss monitoring. This entire process effectively establishes an ensemble structure within a single model, integrating heterogeneous information sources—sentiment analysis (FinBERT), market structure analysis (TVP-VAR), and the original time series (log returns)—to produce a unified output.

Algorithm 3 Hybrid transformer training and predicting ($\mathcal{A}_{\text{Final}}$).

Require: $\mathcal{D}_{\text{Sentiment}}, \mathcal{D}_{\text{Market}}, \mathbf{R}_{\log}$, SeqLen $L = 60$, Epochs E, T_{Stock} Model
Ensure: Predicted Log Returns $\hat{\mathbf{R}}_{\log}$

Model Config: $N = 8$, $h = 10$, $d_{\text{model}} = 320$, Dropout $p = 0.1$, AdamW ($\lambda = 1e-4$)
$\mathcal{D}_{\text{Features}} \leftarrow \text{Merge}(\mathcal{D}_{\text{Sentiment}}, \mathcal{D}_{\text{Market}}, \mathbf{R}_{\log})$
$\mathcal{S}_{\text{Train}} \leftarrow \text{SlidingWindow}(\mathcal{D}_{\text{Features}}, L)$
Initialize T_{Stock} with He initialization
for Epoch $e = 1$ to E **do**
  for batch ($\mathbf{X}, \mathbf{Y}$, IndexID) in DataLoader($\mathcal{S}_{\text{Train}}$) **do**
    Unpack batch $\mathbf{X}$ into $\mathbf{R}_{\log}, M_{k,t}, S_t, C_t$
    $\mathbf{H}_{\text{LSTM}} \leftarrow \text{LSTM}(\mathbf{R}_{\log})$ ▷ Extract temporal features from price
    $\mathbf{X}_{\text{Fused}} \leftarrow \text{Concatenate}(M_{k,t}, S_t, C_t, \mathbf{H}_{\text{LSTM}})$ ▷ Fuse heterogeneous features
    $\mathbf{Z}_{\text{Emb}} \leftarrow \text{LinearProj}(\mathbf{X}_{\text{Fused}}) + \text{IndexEmb}(\text{IndexID}) + \text{PositionalEmb}$
    $\mathbf{Z}_{\text{Enc}} \leftarrow \text{TransformerEncoder}(\mathbf{Z}_{\text{Emb}})$
    $\hat{\mathbf{Y}} \leftarrow \text{FullyConnectedLayer}(\mathbf{Z}_{\text{Enc,CLS}})$
    $\mathcal{L}_{\text{Huber}} \leftarrow \text{SmoothL1}(\hat{\mathbf{Y}}, \mathbf{Y})$
    $\mathcal{L}_{\text{Composite}} \leftarrow \mathcal{L}_{\text{Huber}} + \lambda_{\text{corr}} \cdot (1 - \text{PearsonCor}(\hat{\mathbf{Y}}, \mathbf{Y}))$
    Minimize($\mathcal{L}_{\text{Composite}}$) via Backprop
  end for
  Check Early Stopping based on Validation Loss
end for
for Time t in Test Period **do**
  $\hat{Y}_{k,t+1} \leftarrow T_{\text{Stock}}(\mathbf{X}_{\text{Test},t})$
  $\mathbf{R}_{\log,k,\text{next}} \leftarrow Y_{k,t+1}$
end for
return Predicted Log Returns $\hat{\mathbf{R}}_{\log}$

To visualize the end-to-end data flow described in the algorithms above, the comprehensive architecture of the proposed model is illustrated in Figure 5. The diagram is divided into three parallel input channels on the left, followed by the fusion and prediction modules on the right. The Sentiment Channel processes raw news headlines through FinBERT to generate probability distributions, which are then aggregated into daily sentiment features (S_t, C_t). The Structural Channel transforms log returns (R_{log}) into dynamic spillover indices via the TVP-VAR model; these indices are visualized as heatmaps and subsequently compressed into structural embeddings by a CNN. The Price Channel utilizes an LSTM to extract temporal hidden states (H_{LSTM}) from historical price data. As depicted in the center of the figure, these heterogeneous features are concatenated and linearly projected before being augmented with positional embeddings. Finally, the fused sequence enters the Transformer Encoder stack, where N layers of MHSA and Feed-Forward Networks (FFN) capture long-range dependencies. The final prediction ($\hat{Y}$) is generated by passing the extracted [CLS] token through a Fully Connected (FC) layer.

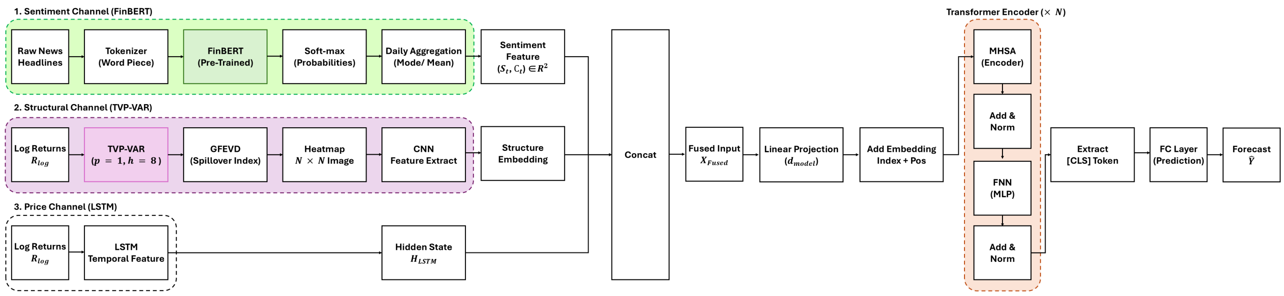


Figure 5. Schematic diagram of the proposed model integrating FinBERT-based sentiment, TVP-VAR-based market structure.

To ensure the reproducibility of the proposed framework and clarify the feature engineering process, we summarized the rigorous pre-processing steps and parameter configurations in Table 8.

Table 8. Detailed specifications of data pre-processing and model parameters for reproducibility.

Module	Parameter / Step	Value / Specification
Sentiment (Text)	Normalization	NFKC, HTML & URL removal
	Filtering Criteria	Exclude headlines with < 3 words
	Label Mapping	Positive (+1), Neutral (0), Negative (−1)
	Daily Aggregation	Sentiment: Mode, Confidence: Mean
Structure (TVP-VAR)	Lag Order (p)	1
	Decay Factors	$\delta = 0.99$, $\lambda = 0.99$
	Covariance Smoothing	$\alpha = 0.97$ (EWMA)
	Forecast Horizon (H)	10 days
Feature Imaging	Image Resolution	128×128 pixels
	Interpolation	Bilinear Interpolation

For the textual data, we applied Normalization Form Compatibility Composition (NFKC) unicode normalization and filtered out headlines with fewer than three words to minimize noise. The sentiment labels derived from FinBERT were mapped to numerical values (+1.0 for Positive, 0.0 for Neutral, and −1.0 for Negative), which were then aggregated daily using the mode for sentiment polarity and the arithmetic mean for confidence scores.

Regarding the structural features, the TVP-VAR model was estimated using a Kalman Filter with decay factors set to $\delta = 0.99$ and $\lambda = 0.99$, alongside a covariance smoothing factor of $\alpha = 0.97$. The resulting spillover indices (forecast horizon $H = 10$) were min-max normalized to the [0, 1] range and upsampled to a resolution of 128×128 pixels via bilinear interpolation to serve as standardized inputs for the CNN feature extractor.

The key advantages of the proposed model are as follows.

First, it enables a multi-dimensional analysis of the market by systematically integrating data from multiple domains. This model overcomes the fundamental limitations of traditional time-series models that rely solely on historical price data. To achieve this, it uses log return data, which captures the intrinsic characteristics of the assets, as its primary input. At the same time, it employs a FinBERT model to extract and integrate market sentiment from financial news headlines as a quantitative indicator, which often provides critical information preceding price fluctuations. Furthermore, through a combination of a TVP-VAR and CNN, the model captures the dynamic connectivity of the entire market system, extending beyond individual assets to incorporate macro-level structural changes into the analysis. By organically fusing these heterogeneous data sources (sentiment, structure, and price), the model provides a multi-faceted analysis of complex market dynamics that cannot be grasped from a single data source, thereby enhancing the robustness and accuracy of its predictions.

Second, the model adopts an innovative approach by utilizing the dynamic changes in market structure as quantitative features. In financial markets, the interdependence between assets is not a fixed value but a dynamic characteristic that constantly evolves in response to external shocks. The proposed model employs a novel methodology that leverages this exact process of structural reorganization as core predictive information. To this end, a TVP-VAR model is used to explicitly estimate the time-varying spillover effects between markets, thereby capturing their dynamic network structure and mutual influence. Subsequently, the estimated network structure at each time point is treated as a structural image, to which a CNN is applied to extract key features. This process quantifies complex, non-linear patterns, such as 'risk transmission structures centered on a specific country' or 'intensified synchronization across the market', going beyond simple correlation analysis. By integrating these systemic drivers - which are imperceptible from price data alone - into the predictive model, its performance is significantly enhanced.

Accordingly, the proposed model is a sophisticated framework that offers a significant advancement for financial market prediction. It effectively extracts the unique characteristics of each data type through specialized modules and integrates them into a powerful time-series learning model, the Transformer.

5. Empirical results

5.1. Hyperparameter settings and implementation

To ensure reproducibility and conduct an objective performance evaluation, this study adhered to the rigorous experimental configurations summarized in Table 9.

Table 9. Experimental configuration and hyperparameter settings for the proposed Hybrid Transformer.

Category	Setting
Data split	Train: 2020–2023, Test: 2024
Sequence length	$L = 60$
Input dimension	134 (CNN 128 + sentiment/conf. 2 + other returns 4)
Scaler	RobustScaler (fit on train)
Transformer	$N = 8$, $h = 10$, $d_{\text{model}} = 320$
Dropout	$p = 0.10$
Optimizer	AdamW
Learning rate	1×10^{-4}
Epochs	500
Input noise	$\sigma = 0.01$
Loss	SmoothL1 (Huber) + correlation regularization

First, to fundamentally eliminate look-ahead bias inherent in time-series forecasting, data splitting was performed chronologically without random shuffling. The dataset from 2020 to 2023 was utilized for model training, while data from January to December 2024 were reserved exclusively for testing to validate out-of-sample prediction performance. The input sequence length (L) was set to 60 days, designed to capture medium-term market trends effectively.

The input dimension comprises a total of 134 dimensions. This includes 128-dimensional market structure embeddings compressed via TVP-VAR and CNN, 2-dimensional sentiment scores and confidence levels extracted by FinBERT, and the 4-dimensional log returns of the target national indices. In the preprocessing stage, a RobustScaler was applied to minimize the impact of outliers frequent in financial time series. Crucially, the scaler statistics were computed solely based on the training data to prevent data leakage.

The architecture of the Transformer was configured with sufficient depth and width to learn the complex non-linearities of the financial market. The encoder consists of 8 layers ($N = 8$), with the number of multi-head attention heads (h) set to 10 and the hidden dimension (d_{model}) to 320. To prevent overfitting and enhance generalization, a dropout rate of 0.1 was applied, and slight noise ($\sigma = 0.01$) was injected into the input data to ensure model robustness. The AdamW optimizer with weight decay was employed, with a learning rate set to 1×10^{-4}. The loss function utilizes SmoothL1 Loss, which is less sensitive to abrupt errors, combined with a regularization term based on Pearson Correlation to induce directional consistency between predicted and actual values.

5.2. Statistical performance evaluation

To assess the impact of stochasticity arising from weight initialization and mini-batch sampling on prediction performance, we conducted a robustness evaluation using multiple random seeds. Across all market indices, the training and evaluation procedures were independently repeated using five distinct random seeds; the corresponding results are summarized in Table 10. As observed in the table, the Mean Squared Error (MSE) of the proposed model exhibits limited variation around the mean across different seeds, accompanied by generally low standard deviations (Std. Dev.). These findings underscore the training stability and reproducibility of the model against random initialization. Furthermore, to ensure consistency in comparative analysis, the main text reports results based on the specific seed configuration used in the initial experiments, while the robustness of these findings is further corroborated by the multi-seed analysis presented in Table 10.

Table 10. Robustness to random initialization: Multi-Seed MSE of the proposed model.

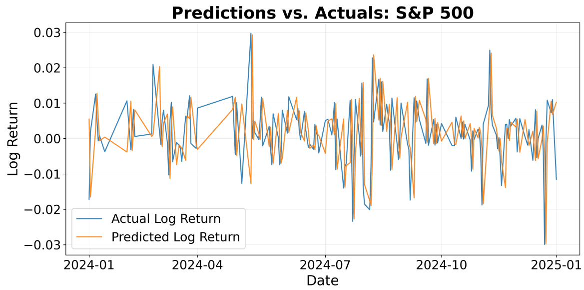
Index	Seed					Statistics	
	1	2	3	4	5	Mean	Std. Dev.
S&P 500	0.000103	0.000103	0.000090	0.000094	0.000109	0.000100	0.000007
FTSE 100	0.000070	0.000052	0.000059	0.000081	0.000090	0.000070	0.000016
CSI 300	0.000252	0.000223	0.000263	0.000278	0.000216	0.000246	0.000026
Nikkei 225	0.000265	0.000211	0.000231	0.000210	0.000271	0.000238	0.000029

Figure 6 visually compares the time series of log returns predicted by the proposed model against the actual log returns for four major stock market indices. Overall, the model's predictions effectively track the direction and volatility of the actual log returns. Notably, even during periods of surging volatility, the predicted values consistently follow the trend of the actuals. This suggests that the model is not merely reverting to a historical average but is meaningfully learning the dynamic shifts in the market.

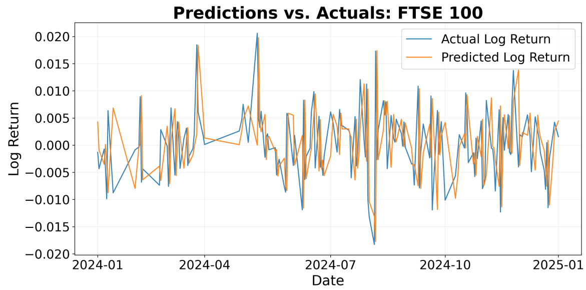
However, a temporary divergence between the predicted and actual values is commonly observed across all indices around April 2024. This discrepancy is analyzed as a combined result of the model's inherent limitations and an unforeseeable macroeconomic shock that impacted the market at the time. Specifically, the U.S. Consumer Price Index (CPI) for March, released in early April 2024, exceeded market expectations, amplifying concerns about a potential delay in interest rate cuts by the Federal Reserve (`https://www.reuters.com/markets/us/us-consumer-prices-rise-more-than-expected-march-2024-04-10/`). Although the proposed model is trained on historical data, news text, and inter-market interaction patterns, this event highlights its limitations in forecasting abrupt shifts in market sentiment that deviate from established patterns.

This CPI shock triggered a stronger dollar and concerns over global capital outflows, simultaneously impacting not only the U.S. market but also the equity markets in the U.K., Japan, and China. Furthermore, the shock exposes an intrinsic limitation of news sentiment analysis models. While models like FinBERT excel at quantifying the textual nuances (e.g., positive, negative, neutral) of published news, they fail to predict the intensity of the market reaction to that information [86, 87]. The language model can measure the negative sentiment level of the news, but the observed adverse market reaction is better interpreted as a consequence of the news increasing uncertainty about the future path of monetary policy. This implies that the market did more than simply price in the negative information; it reassessed asset values by demanding a higher risk premium due to heightened unpredictability [88,89]. When such an asymmetry exists between textual sentiment and the magnitude of the actual market response, predictions based on sentiment analysis can exhibit larger errors.

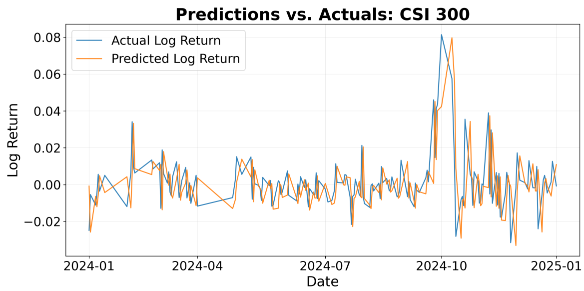
In summary, the visual analysis confirms that the proposed model effectively captures overall market trends while exhibiting clear limitations in the face of unpredictable exogenous shocks. To move beyond this qualitative assessment and to objectively verify the model's predictive accuracy and the contribution of each component, additional analyses were conducted using quantitative evaluation metrics.



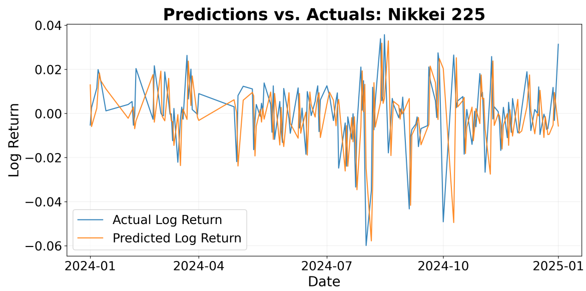
(a) S&P 500 (United States)



(b) FTSE 100 (United Kingdom)



(c) CSI 300 (China)



(d) Nikkei 225 (Japan)

Figure 6. Forecasting performance of the proposed model: A time-series comparison between predicted and actual log returns across four indices.

Tables 11 and 12 present a quantitative evaluation of the predictive performance of the proposed model. The analysis results indicate that the proposed model consistently demonstrates superior prediction accuracy over all comparison models across all stock indices and evaluation metrics (MSE, MAE, and RMSE).

Table 11. Ablation study results illustrating the impact of removing Sentiment and TVP-VAR based market structure features on forecasting accuracy.

Index	Proposed Model			No TVP-VAR			No Sentiment Analysis			Only Log Return Data		
	MSE	MAE	RMSE	MSE	MAE	RMSE	MSE	MAE	RMSE	MSE	MAE	RMSE
S&P 500	0.000103	0.007743	0.010166	0.000124	0.008715	0.011153	0.000124	0.008673	0.011115	0.000121	0.008587	0.011015
FTSE 100	0.000070	0.006331	0.008356	0.000073	0.006618	0.008565	0.000073	0.006548	0.008520	0.000072	0.006570	0.008509
CSI 300	0.000252	0.011897	0.015886	0.000342	0.012275	0.018498	0.000343	0.012270	0.018518	0.000346	0.012303	0.018597
Nikkei 225	0.000265	0.012528	0.016273	0.000306	0.013335	0.017504	0.000295	0.013051	0.017187	0.000303	0.013191	0.017400

Table 12. Performance benchmarking of the proposed framework against established econometric (VAR, GARCH, ARIMA, CIR#) and deep learning (LSTM, N-BEATS) baselines.

Index	VAR			GARCH			ARIMA		
	MSE	MAE	RMSE	MSE	MAE	RMSE	MSE	MAE	RMSE
S&P 500	0.000149	0.009437	0.012213	0.000171	0.010089	0.013064	0.000162	0.009936	0.012730
FTSE 100	0.000112	0.008549	0.010573	0.000106	0.008214	0.010288	0.000126	0.008775	0.011215
CSI 300	0.000282	0.011934	0.016789	0.000324	0.012392	0.017988	0.000276	0.012307	0.016605
Nikkei 225	0.000389	0.015464	0.019715	0.000374	0.014567	0.019333	0.000376	0.015166	0.019395

Index	CIR#			LSTM			N-BEATS		
	MSE	MAE	RMSE	MSE	MAE	RMSE	MSE	MAE	RMSE
S&P 500	0.032792	0.146096	0.181085	0.000148	0.009796	0.012148	0.000111	0.008167	0.010528
FTSE 100	0.004452	0.057490	0.066721	0.000104	0.008106	0.010201	0.000082	0.007262	0.009053
CSI 300	0.063351	0.182399	0.251696	0.000296	0.012004	0.017205	0.000275	0.011548	0.016598
Nikkei 225	0.343276	0.419582	0.585897	0.000388	0.015026	0.019697	0.000272	0.012835	0.016498

First, the effectiveness of the proposed model is clearly demonstrated in Table 11. Notably, for the CSI 300 index, the proposed model exhibited the most significant performance improvement, recording a MSE of 0.000252. This represents an error reduction of approximately 26.3% compared to the 0.000342 MSE of the 'No TVP-VAR' model, which excludes dynamic market structure information. This suggests that modeling the time-varying interconnectivity between markets plays a crucial role in enhancing prediction accuracy.

Furthermore, when compared to the 'No Sentiment Analysis' model, which omits FinBERT-based news sentiment information, the error rate increased across all indices. This serves as evidence that investor sentiment provides critical information for explaining micro-level volatility that is difficult to capture with price data alone. In the case of the FTSE 100 index, while the performance difference resulting from variations in input data composition was the smallest among the indices, the proposed model still demonstrated stable predictive performance by achieving the lowest error across all metrics (MSE 0.000070).

In addition, the enhanced performance of the proposed model over the model that used 'Only Log Return Data' illustrates that the synergistic effect of complementing heterogeneous data contributes to an improvement in predictive power.

Furthermore, Table 12 presents a comprehensive comparison with established baselines. These include traditional econometric models such as VAR, GARCH, ARIMA, and the CIR# model adapted from [90, 91], as well as deep learning models like LSTM and Neural Basis Expansion Analysis for Time Series (N-BEATS).

As presented in Table 12, the superiority of the proposed model is evident when compared not only to conventional time-series models but also to state-of-the-art deep learning approaches.

For instance, in predicting the S&P 500 index, the proposed model achieved an MSE of 0.000103. This result surpasses even N-BEATS (0.000111), which exhibited the most competitive performance among the baselines, thereby validating the precision of our model. Moreover, it demonstrates a statistically significant improvement in performance compared to the widely utilized LSTM model (0.000148) as well as traditional econometric models, including VAR (0.000149) and ARIMA (0.000162).

In contrast, the CIR# model yielded a relatively higher prediction error (MSE of 0.032792 for the S&P 500) compared to other comparison groups. These findings suggest that the proposed framework maintains significantly more stable and robust predictive performance when compared with diverse modeling methodologies.

This trend of superior performance was consistently observed across other stock indices. In the case of the Nikkei 225 index, the proposed model's MSE (0.000265) was lower than that of all traditional benchmark models, further confirming the validity of the methodology proposed in this study.

Thus, the enhanced performance of the proposed model over traditional statistical-based models and deep learning models that rely on a single data source indicates its effectiveness. This stems from the model's organic integration of dynamic market structure analysis, via the TVP-VAR model, with the results from the FinBERT-based sentiment analysis.

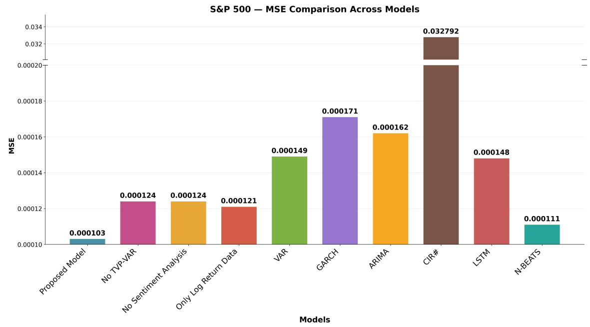
To synthesize these quantitative results and offer a more intuitive understanding of the performance gaps, Figure 7 visually compares the MSE of the proposed model against all benchmark models for each stock index.

The most salient result is that the 'Proposed Model' consistently records the lowest MSE value across all stock indices. This clearly corroborates the quantitative analysis presented earlier in Table 11 and Table 12, visually substantiating the predictive superiority of the proposed model.

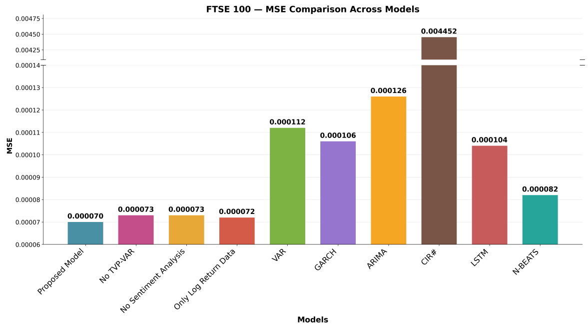
Notably, the MSE of the 'No TVP-VAR' model is discernibly higher than that of the 'Proposed Model' in all markets. This strongly suggests that the dynamic market structure information analyzed through the TVP-VAR model made a crucial contribution to the enhancement of predictive performance. Similarly, the 'No Sentiment Analysis' and 'Only Log Return Data' models also exhibit higher errors than the proposed model, empirically demonstrating that the integration of heterogeneous data can improve a model's predictive power.

A notable observation from the benchmark comparison concerns the performance of the CIR# model, which recorded MSE values distinctly higher than those of other models. This outcome suggests that the traditional econometric framework of CIR# may face challenges in fully capturing the complex, non-linear correlations present in multivariate data.

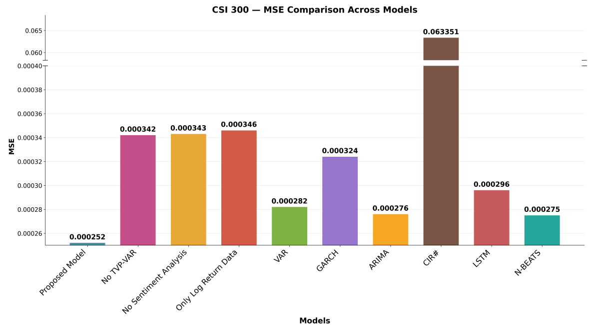
Furthermore, deep learning architectures, such as N-BEATS and LSTM, as well as conventional time-series models like VAR, GARCH, and ARIMA, also exhibited higher error rates compared to the proposed model. While N-BEATS demonstrated improved performance over LSTM in certain indices, it ultimately did not reach the predictive accuracy of the proposed model. These results indicate that existing methodologies relying on single data sources may be less effective in addressing the multifaceted dynamics of financial markets.



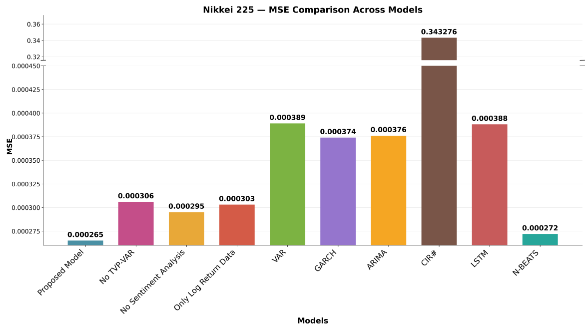
(a) S&P 500 (United States)



(b) FTSE 100 (United Kingdom)



(c) CSI 300 (China)



(d) Nikkei 225 (Japan)

Figure 7. Visual comparison of predictive errors (MSE) demonstrating the superior performance of the proposed model relative to all benchmark models across four major indices.

In summary, Figure 7 serves as powerful visual evidence for our central claim: a multi-modal approach that organically combines heterogeneous data sources achieves a predictive performance unattainable by models that rely on a single source of information.

Furthermore, what is particularly noteworthy about these results is that the enhanced performance of the proposed model is not confined to a specific market but is consistently observed across all stock indices analyzed. This suggests that the proposed multi-modal approach is not overfitted to the unique characteristics of an individual market but possesses a generalized predictive capability effective across diverse market environments.

Moreover, considering the inherent difficulty of securing a predictive edge in financial markets, even a marginal improvement in MSE is a significant outcome. It provides strong evidence that the model successfully leveraged new sources of information that traditional methodologies failed to capture. Therefore, these findings clearly demonstrate that integrating heterogeneous data provides an effective pathway to transcend the fundamental limitations of single-data-source models and to capture complex market dynamics more accurately.

To verify whether the predictive superiority of the proposed model extends beyond simple numerical differences in error metrics (MSE/MAE/RMSE) to statistical significance, we conducted the Diebold-Mariano (DM) test. The DM test evaluates the null hypothesis that the mean difference in the loss series between two competing models is zero. In this study, the loss differential is defined as $(Loss_{ProposedModel} - Loss_{Benchmark})$. Under this definition, a negative DM statistic implies that the proposed model has a significantly smaller average prediction error than the benchmark.

As shown in Table 13, the DM statistics are consistently negative across most comparison models. Notably, in the Ablation Study, the No TVP-VAR model exhibits large negative statistics, such as -5.241 for the S&P 500 and -6.169 for the FTSE 100. This result strongly suggests that the absence of dynamic market structure information significantly degrades predictive performance, thereby validating the effectiveness of integrating TVP-VAR features. Similarly, the 'No Sentiment Analysis' model records a statistic of -3.724 for the FTSE 100, confirming that incorporating news-based investor sentiment is essential for enhancing forecasting accuracy.

Table 13. Diebold–Mariano test statistics comparing the proposed model against baseline models across different indices.

Model	S&P 500	FTSE 100	CSI 300	Nikkei 225
Ablation Studies				
No TVP-VAR	-5.241	-6.169	-4.151	-0.884
No Sentiment Analysis	-1.198	-3.724	-0.027	-0.577
Only Log Return Data	-3.667	-9.001	-3.200	-1.739
Econometric Baselines				
VAR	-0.186	-2.602	-1.060	-0.781
GARCH	-0.403	-3.366	-1.000	-0.799
ARIMA	-0.066	-2.781	-1.004	-0.599
CIR#	-13.877	-13.566	-8.671	-10.335
Deep Learning Baselines				
LSTM	-1.330	-4.365	-0.669	-0.506
N-BEATS	-0.384	-0.386	-0.265	-0.298

The proposed model also demonstrates robust statistical superiority over traditional econometric models and deep learning baselines. It significantly outperforms deep learning models such as LSTM (-1.330 for S&P 500) and N-BEATS (-0.384 for S&P 500). In particular, the comparison with the CIR# model yields extremely large negative values, such as -13.877 for the S&P 500 and -13.566 for the FTSE 100. These values indicate that the univariate CIR# framework fails to sufficiently capture the return fluctuations in the multivariate and non-stationary market environment of this study.

In summary, the DM test results in Table 13 provide statistical evidence that the performance improvements observed in Tables 11 and 12 are not due to chance, but are statistically significant outcomes resulting from the systematic integration of heterogeneous data.

5.3. Economic significance and profitability analysis

In this study, additional analyses were conducted to examine the economic utility of improvements in previously derived statistical prediction errors (e.g., MSE, MAE) regarding investment performance and risk management in actual financial markets. Table 14 was compiled by comprehensively evaluating the Index of Directionality (IDX) proposed by [90], Value at Risk (VaR) and Conditional Value at Risk (CVaR) at the 95% confidence level, and risk-adjusted return metrics, specifically the Sharpe Ratio and Rachev Ratio.

Table 14. Comparison of economic performance metrics across four indices.

Model	IDX	VaR_{95}	$CVaR_{95}$	Sharpe	Rachev
Actual	1.000	−0.014	−0.019	2.081	0.968
Proposed Model	0.506	−0.012	−0.017	1.073	1.238
No TVP-VAR	0.428	−0.012	−0.020	−2.648	0.983
No Sentiment Analysis	0.428	−0.012	−0.020	−2.648	0.983
Only Log Return Data	0.428	−0.012	−0.020	−2.648	0.983
VAR	0.450	−0.012	−0.021	−0.849	0.858
GARCH	0.481	−0.014	−0.018	−0.374	1.079
ARIMA	0.492	−0.017	−0.022	−2.395	0.622
CIR#	0.428	−0.012	−0.020	−2.648	0.983
LSTM	0.502	−0.014	−0.019	0.089	1.043
N−BEATS	0.211	−0.018	−0.023	−1.871	0.193

(a) S&P 500

Model	IDX	VaR_{95}	$CVaR_{95}$	Sharpe	Rachev
Actual	1.000	−0.010	−0.013	0.135	1.124
Proposed Model	0.556	−0.010	−0.012	0.140	1.164
No TVP-VAR	0.500	−0.009	−0.014	−0.993	0.883
No Sentiment Analysis	0.506	−0.009	−0.014	−0.825	0.883
Only Log Return Data	0.494	−0.009	−0.014	−1.014	0.883
VAR	0.486	−0.011	−0.014	−0.986	0.838
GARCH	0.518	−0.008	−0.012	1.736	1.125
ARIMA	0.491	−0.011	−0.015	−2.037	0.787
CIR#	0.475	−0.009	−0.014	−1.214	0.883
LSTM	0.508	−0.011	−0.013	−1.008	1.073
N−BEATS	0.475	−0.011	−0.015	−2.586	0.670

(b) FTSE 100

Model	IDX	VaR_{95}	$CVaR_{95}$	Sharpe	Rachev
Actual	1.000	−0.014	−0.027	0.844	1.565
Proposed Model	0.551	−0.016	−0.026	1.235	1.653
No TVP-VAR	0.494	−0.028	−0.045	−1.652	0.465
No Sentiment Analysis	0.475	−0.028	−0.045	−1.737	0.468
Only Log Return Data	0.506	−0.028	−0.045	−1.513	0.471
VAR	0.481	−0.023	−0.041	−1.745	0.658
GARCH	0.524	−0.015	−0.037	0.334	0.904
ARIMA	0.512	−0.026	−0.041	−0.334	0.706
CIR#	0.489	−0.024	−0.044	−1.620	0.500
LSTM	0.513	−0.015	−0.027	1.080	1.581
N−BEATS	0.493	−0.016	−0.034	−0.456	1.019

(c) CSI 300

Model	IDX	VaR_{95}	$CVaR_{95}$	Sharpe	Rachev
Actual	1.000	−0.022	−0.034	1.156	0.884
Proposed Model	0.519	−0.022	−0.031	0.117	1.077
No TVP-VAR	0.482	−0.021	−0.028	−0.569	1.276
No Sentiment Analysis	0.482	−0.021	−0.028	−0.569	1.276
Only Log Return Data	0.500	−0.021	−0.027	0.075	1.387
VAR	0.506	−0.020	−0.030	1.176	1.162
GARCH	0.508	−0.022	−0.029	0.649	1.260
ARIMA	0.487	−0.019	−0.026	1.154	1.413
CIR#	0.482	−0.021	−0.028	−0.569	1.276
LSTM	0.439	−0.024	−0.034	−0.910	0.927
N−BEATS	0.410	−0.027	−0.039	−5.611	0.369

(d) Nikkei 225

First, examining the IDX results for evaluating directional prediction performance, the proposed hybrid Transformer model showed relatively higher performance compared to most benchmark models and single-data source-based models (Table 14). Notably, in the S&P 500, the proposed model recorded an IDX of 0.506, slightly exceeding 0.5, whereas N-BEATS recorded 0.211, indicating a significant degradation in directional prediction performance. This suggests that models minimizing point prediction errors do not always guarantee directional prediction accuracy, implying a potential discrepancy between error-based performance and direction-based performance. Meanwhile, the proposed model tended to improve directional prediction performance compared to using only simple price time series by integrating dynamic market structure information extracted from TVP-VAR and FinBERT-based investor sentiment information. Indeed, in ablation settings for the S&P 500 where TVP-VAR or sentiment information was removed, the IDX dropped from 0.506 to 0.428, supporting the notion that multimodal information integration can contribute to enhanced directional performance.

Additionally, VaR and CVaR were analyzed to evaluate downside risk and tail risk characteristics. $VaR_{0.95}$ represents the loss quantile at the 95% confidence level, while $CVaR_{0.95}$ measures the average loss (expected value) in the region exceeding that threshold, thereby reflecting extreme loss risks more conservatively. In the S&P 500 results (Panel a of Table 14), the $VaR_{0.95}$ and $CVaR_{0.95}$ of the actual market were −0.014 and −0.019, respectively. The proposed model yielded a $\mathrm{VaR}_{0.95}$ of −0.012 and a $\mathrm{CVaR}_{0.95}$ of −0.017, values relatively close to the actual market's downside risk levels. Conversely, N-BEATS showed a more conservative tail risk estimation with $\mathrm{CVaR}_{0.95}$ of −0.023. This implies that the shapes of the downside tails of predicted return distributions differ by model, and the proposed model tends to produce estimates closer to the actual market's tail risk levels.

Furthermore, the direction and magnitude of risk-adjusted performance generated by each model were compared through Sharpe ratio analysis. In the FTSE 100 market (Panel b of Table 14), the actual market Sharpe ratio was 0.135, and the proposed model was observed at 0.140, showing a risk-return structure similar to the market. In contrast, GARCH recorded a relatively large Sharpe ratio of 1.736. This difference can be interpreted alongside discrepancies in VaR_{95} and CVaR_{95} estimates (e.g., GARCH's VaR_{95} being estimated less conservatively than the actual market), suggesting the possibility that risk-adjusted metrics may be overestimated or underestimated depending on the risk estimation method. Also, in the Nikkei 225 market (Panel d of Table 14), the sign of the Sharpe ratio varied by model. Specifically, negative Sharpe ratios were observed in No TVP-VAR (−0.569), No Sentiment Analysis (−0.569), CIR# (−0.569), LSTM (−0.910), and N-BEATS (−5.611), while positive Sharpe ratios were recorded in VAR (1.176), ARIMA (1.154), GARCH (0.649), Proposed Model (0.117), and Only Log Return Data (0.075). Thus, in the Nikkei 225, this suggests that risk-adjusted performance is not uniformly attributed to specific model groups but can vary according to the risk-return estimation characteristics of each model.

Finally, the Rachev ratio was analyzed to evaluate the asymmetry of the return distribution (Panels a and c of Table 14). The Rachev ratio is a metric measuring the relative magnitude of upside tail gains against downside tail risks; a higher value indicates that performance in the extreme gain region is relatively dominant over the extreme loss region. In the S&P 500, the proposed model recorded 1.238, surpassing both the actual market (0.968) and Only Log Return Data (0.983). Furthermore, in the CSI 300, it showed the highest value among the comparative models in the panel at 1.653, suggesting that the proposed model tends to form a more favorable proportion of upside tail gains relative to downside tail losses. This supports the fact that the proposed model relatively better reflects asymmetric performance characteristics in tail regions, which are difficult to capture with only mean-based performance metrics like the Sharpe ratio.

We comprehensively examined how improvements in error-based prediction performance connect to performance metrics from an economic perspective using IDX, VaR/CVaR, Sharpe ratio, and Rachev ratio. As a result, the proposed model showed a tendency to be relatively superior to comparative models in terms of directional prediction (IDX), and characteristics yielding estimates closer to the actual market's risk levels were observed in downside and tail risk metrics (VaR/CVaR). Moreover, when considering both mean-based risk-adjusted performance (Sharpe) and asymmetric performance in tail regions (Rachev ratio), the proposed model presents potential for improvement in the multifaceted aspects of return-risk characteristics that are hard to capture with a single metric. Furthermore, this suggests that a design integrating macro-market structure (spillover) information and micro-investment sentiment (news data) information can contribute to more realistically reflecting the shape of the prediction distribution and the risk-return structure.

6. Discussion and concluding remarks

This study proposes and empirically validates a novel hybrid Transformer model framework that systematically integrates heterogeneous data to improve the prediction accuracy for volatile and non-linear financial time-series data. To overcome the limitations of prediction models that rely on a single data source, the proposed model incorporates investor sentiment, extracted from financial news data, and the dynamic interaction structure among markets as key predictive variables.

First, we utilized FinBERT, a language model specialized for the financial domain, to convert investor sentiment from news headlines into quantitative indicators of positive, neutral, and negative sentiment. This was intended to incorporate the psychological factors of market participants, which are difficult to capture with price data alone.

Second, a TVP-VAR model was applied to explicitly model the time-varying interconnectedness and spillover effects among national stock indices. The spillover effects derived from the TVP-VAR model were converted into a heatmap format, from which a CNN extracted key structural features to be used as input data.

The extracted log returns, sentiment analysis indicators, and dynamic spillover information were integrated as multi-dimensional input data for the proposed model. This fusion of heterogeneous data plays a pivotal role in enabling the model to learn long-term dependencies in the time series, while simultaneously capturing the complex interactions among various market factors. In predictive experiments on the log returns of four major stock indices (S&P 500, FTSE 100, CSI 300, and Nikkei 225), the proposed model consistently demonstrated superior predictive performance across all evaluation metrics (MSE, MAE, RMSE) when compared to models using a single data source and traditional time-series models (e.g., VAR, LSTM, GARCH, ARIMA, CIR#).

In summary, the findings of this study empirically demonstrate that a multi-modal approach - which considers not only the historical price information of individual assets but also textual data reflecting investor sentiment and the dynamic structural changes of the overall market - can significantly improve the accuracy of financial market predictions. A comparative analysis against benchmark models revealed that the proposed model, utilizing all input data, demonstrated the best performance across all evaluation metrics. This was in contrast to models that used only traditional log-return data or those that excluded either sentiment indicators or market structure information. This clearly indicates that the unique information from each data type acts in a complementary manner, creating a powerful synergy that boosts the robustness and accuracy of the predictive model.

6.1. Contributions and implications

Furthermore, this study empirically substantiates that the dynamic structural change of the market is itself a critical variable containing decisive information for predicting future returns. This is evidenced by the most significant degradation in predictive performance observed when the time-varying interconnectedness among markets was removed from the model. This suggests that the financial market is not merely a collection of individual assets but a complex network where the pathways of capital flow and risk contagion are constantly reshaped in response to external shocks or macroeconomic shifts. The input data generated using the TVP-VAR model in this study effectively captures and learns the non-linear patterns inherent in this dynamic structural reshaping process, leveraging it as a source of predictive power.

Moreover, it was confirmed that sentiment analysis via FinBERT model complements the model's predictive power by effectively quantifying the micro-level market volatility and investors' psychological biases that are not captured by quantitative data. The positive or negative nuances in news headlines often tend to precede price fluctuations, providing important clues that explain the market's short-term overreactions or its sensitivity to uncertainty. Consequently, by simultaneously considering changes in both the macro-level market structure and micro-level investor sentiment, the proposed model is able to comprehensively understand and predict the market's multi-faceted dynamics.

The results of this study carry meaningful implications for both financial market participants and policymakers by highlighting the importance of jointly considering investor sentiment and dynamic market structure in forecasting asset returns. The empirical evidence indicates that price movements are influenced not only by historical return patterns but also by time-varying spillover effects across markets and shifts in investor expectations captured through news sentiment. For practitioners, this suggests that traditional single-source forecasting models may underestimate risk and overlook predictive signals embedded in the evolving structure of market interconnectedness. Incorporating dynamic connectedness measures can improve portfolio allocation, enhance hedging effectiveness, and support more adaptive risk management strategies, particularly during periods of heightened uncertainty when cross-market linkages strengthen.

The proposed framework also provides practical insights for systemic risk monitoring. By transforming TVP-VAR–based spillover measures into structured features, the model captures changes in the transmission of shocks across markets that may serve as early warning indicators of contagion. Financial institutions can leverage such information to complement stress-testing exercises and to better assess exposure to external shocks that originate beyond domestic markets.

From a policy perspective, the findings underscore the relevance of sentiment-driven information and dynamic inter-market linkages for financial stability assessment. Policy announcements, macroeconomic news, and geopolitical developments influence markets not only through fundamentals but also by reshaping investor sentiment and the configuration of market interactions. The results imply that policymakers and central banks could benefit from incorporating sentiment indicators and time-varying connectedness measures into their surveillance frameworks to better anticipate the market impact of policy actions and the potential amplification of shocks. Overall, the study supports a system-oriented view of financial markets and suggests that multi-modal forecasting approaches can contribute to more informed financial decision-making and more effective policy design.

6.2. Limitations

While this study demonstrates significant improvements in stock index forecasting through a hybrid Transformer framework, several limitations remain that merit attention in future research.

First, the geographical scope of the analysis is limited. The empirical analysis of this study was conducted only on four major national indices: the S&P 500 (US), FTSE 100 (UK), CSI 300 (China), and Nikkei 225 (Japan). As such, there may be constraints in generalizing the results to the broader global market, including emerging markets with different levels of market efficiency.

Second, the proposed model is designed to be specialized for forecasting at the macro level of indices; therefore, it has limitations in being directly applied to individual assets. Individual stock prices are influenced not only by common market factors but also substantially by firm-specific factors such as earnings, disclosures, and managerial risks; thus, forecasting at the single-stock level requires additional integration of micro-level variables at the firm level.

Third, there exist issues of data imbalance and selection bias in the news data collection process. In Table 5, cross-country disparities in the volume of articles may induce overfitting or reduce the effective utilization of sentiment signals. In addition, the keyword filtering in Appendix A may lead to the omission of important news and may result in a bias toward large-cap stocks.

6.3. Future studies

To address the identified limitations and further maximize the utility of the proposed hybrid Transformer framework, we suggest three primary directions for future study.

First, future work should test the framework's generalization capability by expanding the target markets beyond four national indices, including emerging markets, to evaluate robustness across different levels of market efficiency and institutional settings.

Second, the framework should be extended beyond macro-level index forecasting toward micro-level prediction for individual assets. To this end, firm-specific news sentiment and within-industry relational networks could be integrated so as to refine the unit of analysis to the stock level. This extension may facilitate the construction of more sophisticated portfolio strategies by jointly accounting for market-wide dynamics and firm-specific risk.

Third, future research should diversify data collection channels to mitigate cross-country information asymmetry and to more comprehensively reflect market-relevant information. To address the limitations of relying on a single source (Reuters) and keyword-based retrieval, expanding the data scope to alternative data sources may help establish a more balanced forecasting environment, even in markets where information is relatively scarce.

Appendix A: Keyword filters for news collection

The following rules were used to collect Reuters headlines by index. Each entry specifies the country context, inclusion `keywords`, and `exclude` terms applied during scraping.

Listing 1. Index-level keyword and exclusion rules used in data collection

```
INDEX_FILTERS = {
    'S&P 500': {
        'country': 'USA',
        'keywords': [
            'u.s.', 'us', 'usa', 'america', 'american', 'united states',
            'wall street', 'dow jones', 'nasdaq', 's&p', 'nyse',
            'federal reserve', 'fed', 'fomc', 'powell', 'yellen',
            'white house', 'congress', 'senate', 'treasury',
            'biden', 'trump', 'democrat', 'republican', 'gop',
            'apple', 'microsoft', 'amazon', 'google', 'meta', 'facebook',
            'tesla', 'nvidia', 'jpmorgan', 'goldman sachs',
            'bank of america', 'wells fargo', 'citigroup',
            'dollar', 'greenback', 'washington'
```

```
        ],
        'exclude': ['canada', 'mexico', 'brazil', 'latin america']
    },
    'FTSE 100': {
        'country': 'UK',
        'keywords': [
            'uk', 'u.k.', 'britain', 'british', 'england', 'london',
            'ftse', 'lse', 'london stock',
            'bank of england', 'boe', 'threadneedle',
            'downing street', 'parliament', 'westminster',
            'sunak', 'rishi', 'starmer', 'labour', 'conservative', 'tory',
            'pound', 'sterling', 'brexit'
        ],
        'exclude': ['ireland', 'scotland only', 'wales only']
    },
    'CSI 300': {
        'country': 'China',
        'keywords': [
            'china', 'chinese', 'beijing', 'shanghai', 'shenzhen',
            'xi jinping', 'xi', 'pboc', 'csrc',
            'yuan', 'renminbi', 'rmb',
            'alibaba', 'tencent', 'baidu', 'huawei', 'xiaomi',
            'bank of china', 'icbc', 'china construction bank',
            'sinopec', 'petrochina', 'china mobile',
        ],
        'exclude': ['japan', 'korea', 'india', 'taiwan', 'hong kong']
    },
    'Nikkei 225': {
        'country': 'Japan',
        'keywords': [
            'japan', 'japanese', 'tokyo', 'nikkei',
            'bank of japan', 'boj', 'kuroda', 'ueda',
            'kishida', 'fumio', 'ldp', 'diet',
            'yen', 'jpy',
            'toyota', 'sony', 'nintendo', 'softbank', 'honda',
            'mitsubishi', 'mitsui', 'sumitomo', 'nissan',
        ],
        'exclude': ['china', 'korea', 'taiwan', 'southeast asia']
    },
}
```

Author contributions

D.-J. Kim and S.-Y. Choi: Conceptualization; D.-J. Kim: Methodology, data curation, visualization; D.-J. Kim, E. Noh, and S.-Y. Choi: Formal analysis, investigation, writing – Original Draft; E. Noh, and S.-Y. Choi: Writing – Review & Editing; S.-Y. Choi: Supervision. All authors have read and approved the final version of the manuscript for publication.

Use of Generative-AI tools declaration

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

Acknowledgments

This work was supported by the National Research Foundation of Korea (NRF) grant funded by the Korea government (MSIT) (No. RS-2024-00454493).

Conflict of interest

Prof. Sun-Yong Choi is the Guest Editor of special issue "Recent Advances in Mathematical Modeling of Financial Economics" for AIMS Mathematics. Prof. Sun-Yong Choi was not involved in the editorial review and the decision to publish this article. All authors declare no conflicts of interest in this paper.

References

1. X. Zhang, Y. Zong, P. Du, S. Wang, J. Wang, Framework for multivariate carbon price forecasting: A novel hybrid model, *J. Environ. Manage.*, **369** (2024), 122275. https://doi.org/10.1016/j.jenvman.2024.122275
2. L. Xie, Z. Chen, S. Yu, Deep convolutional transformer network for stock movement prediction, *Electronics*, **13** (2024), 4225. https://doi.org/10.3390/electronics13214225
3. A. Huang, L. Qiu, Z. Li, Applying deep learning method in TVP-VAR model under systematic financial risk monitoring and early warning, *J. Comput. Appl. Math.*, **382** (2021), 113065. https://doi.org/10.1016/j.cam.2020.113065
4. A. Attarzadeh, M. Isayev, F. Irani, Dynamic interconnectedness and portfolio implications among cryptocurrency, gold, energy, and stock markets: A TVP-VAR approach, *Sustain. Futures*, **8** (2024), 100375. https://doi.org/10.1016/j.sftr.2024.100375
5. J. Liu, H. Zhu, Z. Huang, L. Deng, Dynamic forecasting of exchange rate spillovers with TVP-VAR and deep learning models, *Financ. Res. Lett.*, **86** (2025), 108677. https://doi.org/10.1016/j.frl.2025.108677 https://doi.org/10.1016/j.frl.2025.108677
6. J. Kim, H. S. Kim, S. Y. Choi, Forecasting the S&P 500 index using mathematical-based sentiment analysis and deep learning models: A FinBERT transformer model and LSTM, *Axioms*, **12** (2023), 835. https://doi.org/10.3390/axioms12090835

7. M. A. B. Jeet, R. M. Haque, M. A. I. Sayem, A. Arman, R. M. Rahman, *Using deep learning models and FinBERT to predict the stock price of top banks in the Dhaka Stock Exchange*, In: 2024 IEEE/ACIS 24th International Conference on Computer and Information Science (ICIS), IEEE, 2024, 121–126. https://doi.org/10.1109/ICIS61260.2024.10778363
8. H. D. Jung, B. Jang, Enhancing financial sentiment analysis ability of language model via targeted numerical change-related masking, *IEEE Access*, **12** (2024), 50809–50820. https://doi.org/10.1109/ACCESS.2024.3385855
9. L. Su, X. Zuo, R. Li, X. Wang, H. Zhao, B. Huang, A systematic review for transformer-based long-term series forecasting, *Artif. Intell. Rev.*, **58** (2025), 80. https://doi.org/10.1007/s10462-024-11044-2
10. W. Zhang, J. Huang, R. Wang, C. Wei, W. Huang, Y. Qiao, *Integration of Mamba and Transformer - MAT for long-short range time series forecasting with application to weather dynamics*, In: 2024 International Conference on Electrical, Communication and Computer Engineering (ICECCE), IEEE, 2024, 1–6. https://doi.org/10.1109/ICECCE63537.2024.10823516
11. S. A. Farimani, M. V. Jahan, A. M. Fard, An adaptive multimodal learning model for financial market price prediction, *IEEE Access*, **12** (2024), 121846–121863. https://doi.org/10.1109/ACCESS.2024.3441029
12. P. Chen, Z. Boukouvalas, R. Corizzo, A deep fusion model for stock market prediction with news headlines and time series data, *Neural Comput. Appl.*, **36** (2024), 21229–21271. https://doi.org/10.1007/s00521-024-10303-1
13. S. S. Begum, N. S. Fatima, M. Abbas, *Enhancing stock market prediction with an LSTM-attention model integrating macroeconomic and sentiment features*, In: 2024 International Conference on Innovative Computing, Intelligent Communication and Smart Electrical Systems (ICSES), IEEE, 2024, 1–6. https://doi.org/10.1109/ICSES63760.2024.10910478
14. G. N. Rodrigues, M. N. H. Mir, M. S. M. Bhuiyan, M. Al Rafi, K. N. R. Fuad, M. S. Islam, et al., *MiniBert24: A lightweight transformer-based model for stock market movement prediction*, In: 2024 IEEE 3rd International Conference on Robotics, Automation, Artificial-Intelligence and Internet-of-Things (RAAICON), IEEE, 2024, 293–298.
15. M. B. Mahendran, A. K. Gokul, P. Lakshmi, S. Pavithra, *Comparative advances in financial sentiment analysis: A review of BERT, FinBert, and large language models*, In: 2025 3rd International Conference on Intelligent Data Communication Technologies and Internet of Things (IDCIoT), IEEE, 2025, 39–45. https://doi.org/10.1109/IDCIOT64235.2025.10914764
16. S. Yang, S. Yang, F. Mai, *Financial semantic textual similarity: A new dataset and model*, In: 2024 IEEE Symposium on Computational Intelligence for Financial Engineering and Economics (CIFEr), IEEE, 2024, 1–8. https://doi.org/10.1109/CIFEr62890.2024.10772793
17. Y. Shen, P. K. Zhang, *Financial sentiment analysis on news and reports using large language models and FinBert*, In: 2024 IEEE 6th International Conference on Power, Intelligent Computing and Systems (ICPICS), IEEE, 2024, 717–721. https://doi.org/10.1109/ICPICS62053.2024.10796670
18. O. Shobayo, S. Adeyemi-Longe, O. Popoola, B. Ogunleye, Innovative sentiment analysis and prediction of stock price using FinBERT, GPT-4 and logistic regression: A data-driven approach, *Big Data and Cogn. Comput.*, **8** (2024), 143. https://doi.org/10.3390/bdcc8110143

19. S. B. Priya, M. Kumar, N. P. JD, N. Krithika, *Advanced financial sentiment analysis using FinBERT to explore sentiment dynamics*, In: 2025 3rd International Conference on Intelligent Data Communication Technologies and Internet of Things (IDCIoT), IEEE, 2025, 889–897. https://doi.org/10.1109/IDCIOT64235.2025.10915080
20. D. Gabauer, R. Gupta, On the transmission mechanism of country-specific and international economic uncertainty spillovers: Evidence from a TVP-VAR connectedness decomposition approach, *Econ. Lett.*, **171** (2018), 63–71. https://doi.org/10.1016/j.econlet.2018.07.007
21. K. Nyakurukwa, Y. Seetharam, Cross-country categorical economic policy uncertainty spillovers: Evidence from a conditional connectedness TVP-VAR framework, *J. Finan. Econ. Policy*, **15** (2023), 164–181. https://doi.org/10.1108/JFEP-10-2022-0256
22. L. Zhao, F. Wen, X. Wang, Interaction among China carbon emission trading markets: Nonlinear Granger causality and time-varying effect, *Energ. Econ.*, **91** (2020), 104901. https://doi.org/10.1016/j.eneco.2020.104901
23. B. Lucey, B. Ren, Time-varying tail risk connectedness among sustainability-related products and fossil energy investments, *Energ. Econ.*, **126** (2023), 106812. https://doi.org/10.1016/j.eneco.2023.106812
24. B. Tang, D. S. Matteson, *Probabilistic transformer for time series analysis*, In: Advances in Neural Information Processing Systems, Curran Associates, Inc., **34** (2021), 23592–23608.
25. H. Cao, Z. Huang, T. Yao, J. Wang, H. He, Y. Wang, *Inparformer: Evolutionary decomposition transformers with interactive parallel attention for long-term time series forecasting*, In: Proceedings of the AAAI Conference on Artificial Intelligence, **37** (2023), 6906–6915. https://doi.org/10.1609/aaai.v37i6.25845
26. S. Sridhar, S. Sanagavarapu, *Multi-head self-attention transformer for Dogecoin price prediction*, In: 2021 14th International Conference on Human System Interaction (HSI), IEEE, 2021, 1–6. https://doi.org/10.1109/HSI52170.2021.9538640
27. Y. Hao, L. Dong, F. Wei, K. Xu, *Self-attention attribution: Interpreting information interactions inside transformer*, In: Proceedings of the AAAI Conference on Artificial Intelligence, **35** (2021), 12963–12971. https://doi.org/10.1609/aaai.v35i14.17533
28. Y. Liu, H. Wu, J. Wang, M. Long, *Non-stationary transformers: Exploring the stationarity in time series forecasting*, In: Advances in Neural Information Processing Systems, Curran Associates, Inc., **35** (2022), 9881–9893.
29. H. Zhou, S. Zhang, J. Peng, S. Zhang, J. Li, H. Xiong, et al., *Informer: Beyond efficient transformer for long sequence time-series forecasting*, In: Proceedings of the AAAI conference on artificial intelligence, **35** (2021), 11106–11115. https://doi.org/10.1609/aaai.v35i12.17325
30. R. Gupta, M. Chen, *Sentiment analysis for stock price prediction*, In: 2020 IEEE conference on multimedia information processing and retrieval (MIPR), IEEE, 2020, 213–218. https://doi.org/10.1109/MIPR49039.2020.00051
31. Z. Jin, Y. Yang, Y. Liu, Stock closing price prediction based on sentiment analysis and LSTM, *Neural Comput. Appl.*, **32** (2020), 9713–9729. https://doi.org/10.1007/s00521-019-04504-2
32. X. Li, P. Wu, W. Wang, Incorporating stock prices and news sentiments for stock market prediction: A case of Hong Kong, *Inform. Process. Manag.*, **57** (2020), 102212. https://doi.org/10.1016/j.ipm.2020.102212

33. T. Sidogi, R. Mbuvha, T. Marwala, *Stock price prediction using sentiment analysis*, In: 2021 IEEE International Conference on Systems, Man, and Cybernetics (SMC), IEEE, 2021, 46–51. https://doi.org/10.1109/SMC52423.2021.9659283
34. N. Jing, Z. Wu, H. Wang, A hybrid model integrating deep learning with investor sentiment analysis for stock price prediction, *Expert Syst. Appl.*, **178** (2021), 115019. https://doi.org/10.1016/j.eswa.2021.115019
35. E. Ramos-Pérez, P. J. Alonso-González, J. J. Núñez-Velázquez, Multi-transformer: A new neural network-based architecture for forecasting S&P volatility, *Mathematics*, **9** (2021), 1794. https://doi.org/10.3390/math9151794
36. K. Cao, T. Zhang, J. Huang, Advanced hybrid LSTM-transformer architecture for real-time multi-task prediction in engineering systems, *Sci. Rep.*, **14** (2024), 4890. https://doi.org/10.1038/s41598-024-55483-x
37. S. Dhanasekaran, D. Gopal, J. Logeshwaran, N. Ramya, A. O. Salau, Multi-model traffic forecasting in smart cities using graph neural networks and transformer-based multi-source visual fusion for intelligent transportation management, *Int. J. Intell. Transp.*, **22** (2024), 518–541. https://doi.org/10.1007/s13177-024-00413-4
38. K. Karthika, P. Balasubramanie, S. Harishmitha, P. Shanmugapriya, T. Ramya, *Deep learning based hybrid transformer model for stock price prediction*, In: 2025 International Conference on Multi-Agent Systems for Collaborative Intelligence (ICMSCI), IEEE, 2025, 1603–1608. https://doi.org/10.1109/ICMSCI62561.2025.10894439
39. R. Engle, Dynamic conditional correlation: A simple class of multivariate generalized autoregressive conditional heteroskedasticity models, *J. Bus. Econ. Stat.*, **20** (2002), 339–350. https://doi.org/10.1198/073500102288618487
40. R. Engle, *Anticipating correlations: A new paradigm for risk management*, Princeton University Press, 2009. https://doi.org/10.1515/9781400830190
41. F. X. Diebold, K. Yilmaz, Better to give than to receive: Predictive directional measurement of volatility spillovers, *Int. J. Forecasting*, **28** (2012), 57–66. https://doi.org/10.1016/j.ijforecast.2011.02.006
42. N. Antonakakis, I. Chatziantoniou, D. Gabauer, Refined measures of dynamic connectedness based on time-varying parameter vector autoregressions, *J. Risk Financ. Manag.*, **13** (2020), 84. https://doi.org/10.3390/jrfm13040084
43. N. Antonakakis, R. Kizys, Dynamic spillovers between commodity and currency markets, *Int. Rev. Financ. Anal.*, **41** (2015), 303–319. https://doi.org/10.1016/j.irfa.2015.01.016
44. H. Wang, Z. Xie, D. K. Chiu, K. K. Ho, Multimodal market information fusion for stock price trend prediction in the pharmaceutical sector, *Appl. Intell.*, **55** (2025), 77. https://doi.org/10.1007/s10489-024-05894-0
45. W. An, L. Wang, Y. R. Zeng, Social media-based multi-modal ensemble framework for forecasting soybean futures price, *Comput. Electron. Agr.*, **226** (2024), 109439. https://doi.org/10.1016/j.compag.2024.109439
46. C. Li, Commodity demand forecasting based on multimodal data and recurrent neural networks for e-commerce platforms, *Intell. Syst. Appl.*, **22** (2024), 200364. https://doi.org/10.1016/j.iswa.2024.200364

47. W. Huang, J. Zhao, X. Wang, Model-driven multimodal LSTM-CNN for unbiased structural forecasting of European Union allowances open-high-low-close price, *Energ. Econ.*, **132** (2024), 107459. https://doi.org/10.1016/j.eneco.2024.107459
48. C. Hong, Q. He, Integrating financial knowledge for explainable stock market sentiment analysis via query-guided attention, *Appl. Sci.*, **15** (2025), 6893. https://doi.org/10.3390/app15126893
49. G. Lombardo, A. Bertogalli, S. Consoli, D. R. Recupero, Natural language processing and deep learning for bankruptcy prediction: An end-to-end architecture, *IEEE Access*, **12** (2024), 151075–151091. https://doi.org/10.1109/ACCESS.2024.3476621
50. X. Ding, Y. Zhang, T. Liu, J. Duan, *Deep learning for event-driven stock prediction*, In: Proceedings of the 24th International Conference on Artificial Intelligence, AAAI Press, 2015, 2327–2333.
51. K. Koh, The impact of investor sentiment on stock returns in developed and emerging markets: The case of US and South Korea, *Emerg. Mark. Financ. Tr.*, 2025, 1–24. https://doi.org/10.1080/1540496X.2025.2588236
52. N. A. Boamah, Integration, investor protection rules and global informational inefficiency of emerging financial markets, *SN Bus. Econ.*, **1** (2021), 82. https://doi.org/10.1007/s43546-021-00084-3
53. M. E. Hoque, M. A. S. Zaidi, Global and country-specific geopolitical risk uncertainty and stock return of fragile emerging economies, *Borsa Istanb. Rev.*, **20** (2020), 197–213. https://doi.org/10.1016/j.bir.2020.05.001
54. A. Sehl, M. Eder, News personalization and public service media: The audience perspective in three European countries, *J. Media*, **4** (2023), 322–338. https://doi.org/10.3390/journalmedia4010022
55. P. C. Tetlock, Giving content to investor sentiment: The role of media in the stock market, *J. Finance*, **62** (2007), 1139–1168. https://doi.org/10.1111/j.1540-6261.2007.01232.x
56. G. Birz, E. Devos, S. Dutta, D. Tsang, Is good news good and bad news bad in the REIT market?, *J. Real Estate Portfolio Manag.*, **27** (2021), 43–62. https://doi.org/10.1080/10835547.2021.1967689
57. T. Loughran, B. McDonald, When is a liability not a liability? textual analysis, dictionaries, and 10-ks, *J. Finance*, **66** (2011), 35–65. https://doi.org/10.1111/j.1540-6261.2010.01625.x
58. C. Kearney, S. Liu, Textual sentiment in finance: A survey of methods and models, *Int. Rev. Financ. Anal.*, **33** (2014), 171–185. https://doi.org/10.1016/j.irfa.2014.02.006
59. J. W. Kang, S. Y. Choi, Comparative investigation of GPT and FinBERT's sentiment analysis performance in news across different sectors, *Electronics*, **14** (2025), 1090. https://doi.org/10.3390/electronics14061090
60. J. P. Venugopal, A. A. V. Subramanian, G. Sundaram, M. Rivera, P. Wheeler, A comprehensive approach to bias mitigation for sentiment analysis of social media data, *Appl. Sci.*, **14** (2024), 11471. https://doi.org/10.3390/app142311471
61. I. Ghosh, T. D. Chaudhuri, S. Sarkar, S. Mukhopadhyay, A. Roy, Macroeconomic shocks, market uncertainty and speculative bubbles: A decomposition-based predictive model of Indian stock markets, *China Financ. Rev. Int.*, **15** (2025), 166–201. https://doi.org/10.1108/CFRI-09-2023-0237
62. M. L. Narayana, A. J. Kartha, A. K. Mandal, A. Suresh, A. C. Jose, Ensemble time series models for stock price prediction and portfolio optimization with sentiment analysis, *J. Intell. Inform. Syst.*, **63** (2025), 1079–1103. https://doi.org/10.1007/s10844-025-00928-6

63. M. N. H. Mir, M. S. M. Bhuiyan, M. Al Rafi, G. N. Rodrigues, M. Mridha, M. A. Hamid, et al., Joint topic-emotion modeling in financial texts: A novel approach to investor sentiment and market trends, *IEEE Access*, **13** (2025), 28664–28677. https://doi.org/10.1109/ACCESS.2025.3538760
64. R. Liu, J. Liang, H. Chen, Y. Hu, Analyst reports and stock performance: Evidence from the Chinese market, *Asia-Pac. Financ. Mark.*, 2025, 1–26. https://doi.org/10.1007/s10690-025-09522-w
65. X. L. Gong, H. Y. Ning, X. Xiong, Research on the cross-contagion between international stock markets and geopolitical risks: The two-layer network perspective, *Financ. Innov.*, **11** (2025), 23. https://doi.org/10.1186/s40854-024-00687-3
66. S. D. Yeboah, S. K. Agyei, D. Korsah, M. P. Fumey, P. K. Akorsu, V. Adela, Geopolitical risk and exchange rate dynamics in Sub-Saharan Africa's emerging economies, *Futur. Bus. J.*, **11** (2025), 78. https://doi.org/10.1186/s43093-025-00505-x
67. Y. Ding, X. Zhu, R. Pan, B. Zhang, Network vector autoregression with time-varying nodal influence, *Comput. Econ.*, **66** (2025), 4161–4187. https://doi.org/10.1007/s10614-024-10841-9
68. Q. Liu, C. Xu, Z. Xu, The impact of geopolitical risk on food prices: Evidence from the TVP-SV-VAR model, *J. Econ. Financ.*, **49** (2025), 450–473. https://doi.org/10.1007/s12197-025-09710-4
69. H. Le, A. Bouteska, T. Sharif, M. Z. Abedin, Dynamic interlinkages between carbon risk and volatility of green and renewable energy: A TVP-VAR analysis, *Res. Int. Bus. Financ.*, **69** (2024), 102278. https://doi.org/10.1016/j.ribaf.2024.102278
70. M. Yousfi, H. Bouzgarrou, Exploring interconnections and risk evaluation of green equities and bonds: Fresh perspectives from TVP-VAR model and wavelet-based VaR analysis, *China Financ. Rev. Int.*, **15** (2025), 117–139. https://doi.org/10.1108/CFRI-05-2024-0237
71. F. Ahmadian-Yazdi, A. Sokhanvar, S. Roudari, A. K. Tiwari, Dynamics of the relationship between stock markets and exchange rates during quantitative easing and tightening, *Financ. Innov.*, **11** (2025), 51. https://doi.org/10.1186/s40854-024-00694-4
72. H. G. Souto, A. Moradi, Can transformers transform financial forecasting?, *China Financ. Rev. Int.*, 2024. https://doi.org/10.1108/CFRI-01-2024-0032
73. J. Yang, P. Li, Y. Cui, X. Han, M. Zhou, Multi-sensor temporal fusion transformer for stock performance prediction: An adaptive Sharpe ratio approach, *Sensors*, **25** (2025), 976. https://doi.org/10.3390/s25030976
74. M. Mohseni, I. Mohammadzaman, Robust meta-reinforcement learning for autonomous spacecraft rendezvous with transformer networks under delayed observations, *Int. J. Aeronaut. Space Sci.*, **26** (2025), 2724–2749. https://doi.org/10.1007/s42405-025-00917-7
75. D. Xiang, W. Yang, Z. Zhou, J. Zhang, J. Li, J. Ouyang, et al., DPMFformer: An underwater image enhancement network based on deep pooling and multi-scale fusion transformer, *Earth Sci. Inform.*, **18** (2025), 61. https://doi.org/10.1007/s12145-024-01573-3
76. S. Li, Y. Sun, W. Yue, M. Yao, Y. Han, G. Gui, et al., A novel multi-scale time fusion transformer for long-range spectrum occupancy prediction, *IEEE T. Veh. Technol.*, **74** (2025), 9299–9312. https://doi.org/10.1109/TVT.2025.3540920
77. S. Liu, X. Wang, An improved transformer based traffic flow prediction model, *Sci. Rep.*, **15** (2025), 8284. https://doi.org/10.1038/s41598-025-92425-7

78. G. Campisi, S. Muzzioli, A. Zaffaroni, Nonlinear dynamics in asset pricing: The role of a sentiment index, *Nonlinear Dynam.*, **105** (2021), 2509–2523. https://doi.org/10.1007/s11071-021-06724-5
79. G. Tedeschi, F. Caccioli, M. C. Recchioni, Taming financial systemic risk: Models, instruments and early warning indicators, *J. Econ. Interact. Coor.*, **15** (2020), 1–7. https://doi.org/10.1007/s11403-019-00278-x
80. P. Malo, A. Sinha, P. Korhonen, J. Wallenius, P. Takala, Good debt or bad debt: Detecting semantic orientations in economic texts, *J. Assoc. Inf. Sci. Tech.*, **65** (2014), 782–796. https://doi.org/10.1002/asi.23062
81. D. Araci, FinBERT: Financial sentiment analysis with pre-trained language models, *arXiv preprint*, 2019. https://doi.org/10.48550/arXiv.1908.10063
82. Z. Wang, T. Oates, *Encoding time series as images for visual inspection and classification using tiled convolutional neural networks*, In: Workshops at the twenty-ninth AAAI conference on artificial intelligence, vol. 1. AAAI, 2015, 000091–000096.
83. O. B. Sezer, A. M. Ozbayoglu, Algorithmic financial trading with deep convolutional neural networks: Time series to image conversion approach, *Appl. Soft Comput.*, **70** (2018), 525–538. https://doi.org/10.1016/j.asoc.2018.04.024
84. E. Hoseinzade, S. Haratizadeh, CNNpred: CNN-based stock market prediction using a diverse set of variables, *Expert Syst. Appl.*, **129** (2019), 273–285. https://doi.org/10.1016/j.eswa.2019.03.029
85. S. Barra, S. M. Carta, A. Corriga, A. S. Podda, D. R. Recupero, Deep learning and time series-to-image encoding for financial forecasting, *IEEE/CAA J. Automatic.*, **7** (2020), 683–692. https://doi.org/10.1109/JAS.2020.1003132
86. M. Davidovic, J. McCleary, News sentiment and stock market dynamics: A machine learning investigation, *J. Risk Financ. Manag.*, **18** (2025), 412. https://doi.org/10.3390/jrfm18080412
87. G. Shashank, G. Pandey, S. G. Koolagudi, *Time based sentiment analysis of financial headlines using recurrent neural network*, In: 2025 International Conference on Artificial Intelligence and Data Engineering (AIDE), IEEE, 2025, 752–756. https://doi.org/10.1109/AIDE64228.2025.10987536
88. S. R. Baker, N. Bloom, S. J. Davis, Measuring economic policy uncertainty, *Q. J. Econ.*, **131** (2016), 1593–1636. https://doi.org/10.1093/qje/qjw024
89. L. Pastor, P. Veronesi, Uncertainty about government policy and stock prices, *J. Finance*, **67** (2012), 1219–1264. https://doi.org/10.1111/j.1540-6261.2012.01746.x
90. G. Orlando, M. Bufalo, Interest rates forecasting: Between Hull and White and the CIR#-how to make a single-factor model work, *J. Forecasting*, **40** (2021), 1566–1580. https://doi.org/10.1002/for.2783
91. G. Orlando, M. Bufalo, Time series forecasting with the CIR# model: From hectic markets sentiments to regular seasonal tourism, *Technol. Econ. Dev. Eco.*, **29** (2023), 1216–1238. https://doi.org/10.3846/tede.2023.19294

AIMS Press