VIRTUAL ECONOMICS

ISSN 2657-4047 *(online)*

*Research Article*

# Volatility-Sensitive Forecasting: Deep Learning Model Robustness across Calm and Crisis Periods

*Laszlo Vancsura, Tibor Tatay, and Tibor Bareith*

**Abstract.** This study investigates how the forecasting performance of deep learning models is affected by changing market conditions, with particular emphasis on periods of differing volatility. While deep learning methods are increasingly applied in financial forecasting, relatively limited evidence exists on how their predictive accuracy varies across distinct market regimes. The aim of this paper is to examine whether model performance systematically deteriorates during more turbulent periods and whether this effect differs across model architectures. The empirical analysis covers multiple asset classes, including equities, commodities, cryptocurrencies, and currency pairs, over three partially overlapping periods (2016–2018, 2018–2020, and 2020–2022), capturing both relatively stable and highly volatile market environments. Forecasting performance is evaluated using several recurrent neural network architectures, and the relationship between market volatility and prediction error is analysed using cross-sectional regression techniques. Additional sensitivity analyses are conducted to assess the robustness of the results across alternative asset samples. The results reveal a consistent positive relationship between market volatility and forecasting error, indicating that predictive performance tends to deteriorate under more turbulent market conditions. However, the magnitude of this effect varies across model architectures. In particular, GRU-based architectures appear less sensitive to volatility, whereas the use of multivariate inputs does not consistently improve forecasting accuracy. These findings suggest that greater model complexity does not necessarily translate into more robust performance under changing market conditions. The study is subject to several limitations, including the relatively small cross-sectional sample and the use of standardised model configurations. Future research could extend the analysis by considering alternative model classes, broader datasets, and more comprehensive robustness frameworks.

**Keywords:** STL-based volatility, deep learning, forecasting performance, COVID-19, Russian–Ukrainian conflict.

**Author(s):**

**László Vancsura**
Department of Marketing and Supply Chain Analytics; Institute of Agricultural and Food Economics; Hungarian University of Agriculture and Life Sciences; 7400 Kaposvár, Hungary
E-mail: vancsura.laszlo@uni-mate.hu
https://orcid.org/0000-0002-6345-8209

**Tibor Tatay**
Department of Statistics, Finances and Controlling; Széchenyi István University; 9026 Győr, Hungary
E-mail: tatay@sze.hu
https://orcid.org/0000-0001-5436-4545

**Tibor Bareith**
ELTE, Centre for Economic and Regional Studies, Institute of Economics, 1097 Budapest, Hungary
E-mail: bareith.tibor@krtk.elte.hu
John von Neumann University, Economic Geography and Urban Marketing Centre, 1117 Budapest, Hungary
https://orcid.org/0000-0002-9971-9597

Corresponding author: Tibor Tatay (tatay@sze.hu).

**How to Cite This Article:** Vancsura, L., Tatay, T., & Bareith, T. (2026). Volatility-Sensitive Forecasting: Deep Learning Model Robustness across Calm and Crisis Periods. *Virtual Economics, 9*(1), 28–64. https://doi.org/10.34021/ve.2026.09.01(2)

Received: 9 December 2025. Revised: 6 February 2026. Accepted: 12 March 2026.

## 1. Introduction

The functioning of financial markets is fundamentally shaped by the time-varying nature of volatility. Nonlinear price dynamics characterized by sudden regime shifts make forecasting particularly challenging, as most traditional statistical models assume stationarity that is, the statistical properties of the analysed time series remain constant over time [1]. However, this assumption rarely holds in real financial processes. Over the past decade, global markets have been exposed to extreme shocks, such as the COVID-19 pandemic in 2020 and the Russian invasion of Ukraine in 2022, both of which triggered drastic price fluctuations within a short period [2]. Consequently, both researchers and practitioners have increasingly turned toward methodological approaches capable of flexibly addressing non-stationarity and turbulent environments. The emergence of machine learning and particularly deep learning in time series analysis has opened new avenues for financial forecasting [3-5]. Recurrent neural networks (RNNs), long short-term memory networks (LSTMs), and the more computationally efficient gated recurrent unit (GRU) models are architectures capable of capturing deeper nonlinear patterns in sequential data. However, an increasing body of literature highlights that the performance of these models strongly depends on the level of environmental volatility: under high-volatility conditions, prediction errors can increase substantially [6,7]. In the literature, this issue has been approached from two main directions. One research stream proposes integrating regime-switching models into neural network architectures, such as hybrid HMM–LSTM or HMM–GRU frameworks, which are able to detect hidden transitions between market states [8]. Another direction focuses on attention mechanisms and transformer-based models, aiming to dynamically concentrate the forecasting system on the most relevant patterns, particularly during volatile periods [9-11]. Nevertheless, the empirical investigation of the relationship between volatility and forecasting errors remains incomplete and often relies on inconsistent methodological frameworks. Most studies measure volatility using absolute indicators, which are difficult to compare across asset classes with different scales. In this research, relative volatility is therefore employed, calculated from the residual components of time series that have been detrended and deseasonalized using the Seasonal-Trend Decomposition using Loess (STL) method. The resulting coefficient of variation provides a standardized and comparable basis for understanding how volatility affects predictive performance. The empirical analysis covers twelve financial instruments including stock indices, commodity assets, currency pairs, and cryptocurrencies representing a broad spectrum in terms of both volatility levels and market structures. The analysis distinguishes three periods: a relatively calm phase (2016–2018) and two crisis periods (2020 and 2022). This design allows not only the examination of the relationship between volatility and forecasting performance, but also the assessment of how well different models maintain their robustness under extreme market conditions. A central research question of the study is the extent to which different deep learning models are sensitive to increases in volatility, and whether hybrid architectures are indeed more stable than simpler models. Investigating the relationship between volatility and forecasting errors is not only theoretically important. Risk management and portfolio optimisation in financial institutions rely heavily on short-term forecasts, the accuracy of which may deteriorate dramatically during volatile periods. The practical contribution of this research therefore lies in highlighting that model selection and input structure such as univariate versus multivariate designs, regime information, and attention mechanisms must be carefully considered, particularly when the objective is to develop forecasting systems capable of operating reliably in highly volatile market environments.

While previous literature has primarily focused on improving the predictive accuracy of deep learning models, this study is positioned at the intersection of the Adaptive Markets Hypothesis (AMH) (Lo, 2004) [12] and the bias–variance trade-off known from statistical learning theory [13]. Specifically, it is investigated how the sudden decline in the informational efficiency of markets during crises such as COVID-19 and geopolitical shocks which typically manifests in extreme volatility, affects the robustness of predictive models. It is hypothesized hypothesize that more complex hybrid architectures become vulnerable to overfitting under high noise conditions, whereas simpler network mechanisms (such as GRU models) exhibit greater stability during regime shifts in accordance with the principle of Occam's razor [13].

The study is structured around the following three main research questions (RQs):

RQ1 (Volatility sensitivity): Does trend- and seasonality-adjusted relative volatility (CV_STL) systematically and positively increase forecasting error (MAPE) across different asset classes?

RQ2 (Architectural robustness): Do gated recurrent models (GRU/LSTM) exhibit lower volatility sensitivity during crisis regimes compared to simpler recurrent structures (RNN) and more complex hybrid variants?

RQ3 (Effect of the information set): Does expanding the information set from a univariate input (closing price) to a multivariate input (OHLC) reduce or instead increase volatility sensitivity, and how consistent is this effect across the analysed financial instruments?

By addressing these questions, the study contributes to the literature on financial time series analysis and deep learning in three key ways: (i) Robustness-centred evaluation framework. Rather than focusing solely on traditional point forecast accuracy, the analysis places model robustness and volatility sensitivity at the centre of the evaluation framework. This perspective is particularly critical for risk-sensitive decision-making environments. (ii) Systematic regime-dependent performance evaluation. The predictive performance of deep learning models is examined across multiple asset classes (equities, currencies, commodities, and cryptocurrencies) allowing us to assess how forecasting accuracy responds to both stable periods and structural breaks caused by exogenous shocks such as COVID-19 and geopolitical conflicts. (iii) Application of a unified volatility proxy. The study introduces and applies a standardized measure of relative volatility the coefficient of variation derived from the residual component of the Seasonal-Trend Decomposition using Loess (STL) decomposition which enables consistent comparisons across heterogeneous financial assets and provides an interpretable basis for explaining differences in algorithmic performance.

The remainder of this paper is organized as follows. Section 2 provides a comprehensive literature review, focusing on the challenges of non-stationarity, as well as exogenous regime-shifting and endogenous adaptation mechanisms. Section 3 details the data and methodology, presenting the twelve financial instruments, the three analysis periods, and the deep learning architectures employed, including RNN, LSTM, GRU models, and their hybrid variants. This section also describes the MAPE indicator used for model evaluation and the methodology for STL-based relative volatility calculation. Section 4 presents the empirical results, with a particular focus on the relationship between volatility and forecasting error, and the robustness of different model structures during crisis periods. Section 5 provides the discussion, interpreting the findings in the context of the Adaptive Markets Hypothesis (AMH) and the bias–variance trade-off. Finally, Section 6 offers concluding remarks, highlighting the importance of model selection in turbulent market environments and suggesting directions for future research. For ease of reference, a complete list of the abbreviations used throughout the paper is provided in Appendix A.

## 2. Literature Review

The existing literature on financial time series forecasting under volatile conditions can be broadly structured around three unresolved issues: (i) the measurement of volatility across heterogeneous assets, (ii) the comparative robustness of neural architectures under regime instability, and (iii) the implications of expanding the informational input space. First, with respect to volatility measurement, prior studies typically rely on absolute indicators, which limits cross-asset comparability. This study departs from that convention by employing a relative volatility measure derived from STL residuals, thereby positioning itself within a smaller but growing strand of literature that seeks scale-invariant volatility proxies. Second, the question of architectural robustness remains contested. While hybrid and high-capacity models are frequently shown to outperform simpler alternatives under stable conditions, their behaviour under regime shifts is less clear. Existing research suggests that increased model flexibility may come at the cost of stability, particularly when noise levels rise. In this respect, the present study contributes by explicitly comparing recurrent architectures under varying volatility regimes, rather than assuming performance generalisability. Third, the role of input dimensionality has received comparatively limited attention. Although multivariate inputs are generally assumed to enhance predictive performance, there is emerging evidence that additional features may degrade generalisation when underlying correlation structures become unstable. This study therefore examines whether richer input structures improve or impair forecasting robustness during crisis periods. In contrast to studies that focus on regime-switching, attention mechanisms, or transformer-based models, the present analysis deliberately adopts a more controlled comparative framework centred on recurrent architectures. This choice allows for a clearer identification of the interaction between volatility and model behaviour, thereby addressing a gap in the literature where robustness is often inferred rather than systematically evaluated. Overall, the contribution of this study lies not in proposing a new modelling technique, but in providing a structured empirical comparison that links volatility dynamics to forecasting performance across models and asset classes.

### *2.1. The Challenge of Non-stationarity and the Role of Volatility*

One of the main challenges in the analysis of financial time series is their non-stationary nature. This dynamic behaviour typically manifests in alternating periods of low and high volatility, as well as stable and turbulent market conditions, which pose significant difficulties for traditional static forecasting models. These approaches assume time-invariant parameters and are therefore unable to adequately adapt to sudden changes, which can reduce both predictive accuracy and robustness. Over the past decade, however, machine learning (ML) and deep learning (DL) methods have revolutionized time series forecasting, opening new possibilities for explicitly incorporating volatility into predictive models. The focus of the contemporary literature is increasingly directed toward how forecasting systems can adaptively handle volatile environments while providing more stable and accurate predictions across different market regimes. These approaches can generally be grouped into three main directions: exogenous regime-switching mechanisms, attention-based endogenous selection, and the use of objective functions designed to maximize decision utility.

### 2.2. Adaptation in Exogenous Regime-switching and Hybrid Frameworks

Regime-switching approaches describe price dynamics using hidden, discrete market states in which model parameters vary across regimes. Zhang and Zhang (2022) [14] incorporated Markov regime-switching into an option pricing framework, demonstrating that this statistical method performs better than classical constant-parameter assumptions when modelling volatility clustering and fat-tailed returns. Similarly, Carpinteyro et al. (2021) [15] confirmed the flexibility of Markov structures in capturing asymmetric distributions and structural breaks in precious metal return series. However, a critical issue of these statistical approaches lies in model specification. The subjective choice of the number of regimes and transition probabilities largely determines the resulting model behaviour and often leads to overparameterization. This concern is also emphasized by Shi (2023) [16], who points out that genuine long-term memory and persistence generated by regime switching can be empirically difficult to distinguish. As a result, parameter stability can only be maintained under strictly regime-conditioned modelling.

To overcome the limitations of these statistical models, hybrid solutions have become increasingly common in the field of deep learning. In such systems, a Hidden Markov Model (HMM) compresses hidden states as a feature extraction mechanism, while recurrent neural networks such as RNN, LSTM, and GRU learn nonlinear relationships and longer temporal dependencies. Avinash et al. (2024) [17] and Sivakumar (2025) [8] both demonstrated that under turbulent market conditions these hybrid HMM-DL architectures significantly improve forecasting accuracy compared to baseline models. However, the increased complexity of such models involves important trade-offs. A substantial portion of explanatory power shifts into the black-box layer, and if the regime labels generated by the HMM contain noise, the neural network may internalize that noise as well. In hybrid systems another common side effect is excessive smoothing, where the variance of predictions decreases but part of the informative signal is also lost.

To address the weaknesses of the HMM framework, Aydogan-Kilic and Selcuk-Kestel (2023) [18] proposed a modified hybrid model in which the focus of the loss function is shifted from maximum likelihood estimation toward directly minimizing price deviations. Although this approach establishes a clear objective and increases predictive usefulness, it may compromise the generative coherence of the model. Another purely data-driven direction is proposed by Mari and Mari (2023, 2025) [19,20], who treat regime switching as an endogenous mechanism based on the Softmax prediction probabilities of a deep neural network instead of relying on a classical Markov chain. This approach shifts the focus from point forecasting toward the estimation of conditional distributions, providing greater flexibility in handling volatility. At the same time, however, it significantly reduces interpretability.

### 2.3. Endogenous Adaptation: Attention Mechanisms and Objective Function Optimisation

Another major line of approaches that aim to handle volatility adaptively, without relying on discrete regime definitions, is based on attention mechanisms. These models learn dynamic weighting schemes that selectively emphasize past segments of the time series that are particularly important for forecasting, such as periods surrounding market shocks. Ouyang et al. (2021) [21] developed an attention-LSTM model for analysing systematic risk, distributing relevance across heterogeneous sources, including traditional financial indicators and online sentiment measures. However, the main limitation of these systems often lies not in the model architecture itself but in the measurement noise present in external sentiment data. In contrast, the hybrid architectures proposed by Zhang et al. (2023) [22] and Luo et al. (2024) [23], such as CNN-BiLSTM-Attention and CLATT, apply attention within a closed modelling framework, focusing directly on internal hidden states. These algorithms demonstrate strong and stable performance even during periods of elevated volatility. Nevertheless, it remains an open question how such models respond to deep structural breaks, when volatility increases and the fundamental logic of the market itself undergoes substantial reorganization.

With the growing adoption of transformer-based models, the reinterpretation of training objective functions has also become increasingly important. Yu (2025) [24], in the Temporal Fusion Transformer model, highlights that during highly volatile periods, forecasts optimized solely for error minimization (e.g., mean squared error) may not align with the utility required for investment decision-making. To address this issue, the Sharpe ratio was directly incorporated as a training objective, combining attention mechanisms with risk-sensitive optimisation. In a related contribution, Li and Xu (2024) [25] introduced an attention-based network augmented with a generative adversarial component. Their results confirm the importance of explicitly controlling risk dimensions. While the model demonstrates strong capability in capturing trends, some evaluation metrics indicate a tendency to underestimate systematic risk.

The reviewed literature clearly suggests that forecasting systems must adapt to volatile environments characterized by regime shifts. The key difference among the proposed approaches lies in what they treat as the primary modelling objective. Markov and Hidden Markov Model based frameworks are strong in describing discrete and interpretable market states, but they are highly sensitive to specification choices and to the stability of regime boundaries. Hybrid deep learning and HMM models enhance predictive power by capturing nonlinear relationships, although robustness increasingly depends on the complex neural component. Attention-based and transformer architectures provide flexible endogenous adaptation without requiring explicit regime labels. However, the choice of objective function becomes crucial in determining whether the model ultimately learns statistical accuracy alone or develops predictions that provide meaningful decision utility.

## 3. Data and Methods

### *3.1. Data*

Empirical analysis utilizes daily price observations for a diversified portfolio of financial instruments, encompassing major equity indices (S&P500, DAX, Nikkei225), commodities (crude oil, gold, silver), cryptocurrencies (Bitcoin, Ethereum, Litecoin), and key currency pairs (EUR/USD, GBP/USD, AUD/USD). The total observation window spans from January 1, 2016, to June 30, 2022. To ensure consistency within asset classes, we employed futures prices for crude oil, gold, and silver, while spot prices were used for all other instruments.

The selection of this specific timeframe was motivated by two primary factors. First, it allows for the evaluation of models across distinct market environments: a "calm" baseline (2018), the global pandemic shock (2020), and the geopolitical instability of 2022. Second, this window aligns with the maturation of the cryptocurrency market, providing a sufficient historical record for meaningful comparison with traditional assets. Consequently, this design facilitates a robust assessment of predictive performance. Data were primarily retrieved from Yahoo Finance [26], with cryptocurrency series sourced specifically from CoinMarketCap [27].

During the preprocessing stage, we addressed the challenge of missing values, which was particularly critical for the multi-asset correlation analysis due to differing international market calendars. We resolved these discrepancies using linear interpolation. For the individual forecasting tasks, missing data were handled separately for each product to maintain independence. While the methodology for trend and seasonal decomposition is detailed in the volatility analysis section, the forecasting models were trained without explicit seasonal adjustments; instead, we relied on the models' inherent pattern recognition capabilities to internalize these dynamics. To test the robustness of the deep learning architectures under varying conditions, the dataset was partitioned into three partially overlapping sub-periods: (i) Stable Regime: January 1, 2016 – June 30, 2018; (ii) Pandemic Regime: January 1, 2018 – June 30, 2020; (iii) Conflict Regime: January 1, 2020 – June 30, 2022.

As shown in Table 1, the first period was characterized by steady equity growth and relatively stable volatility levels. While the S&P 500 recorded moderate price dispersion, the European (DAX) and Asian (Nikkei225) benchmarks displayed higher standard deviations, indicating more significant price swings in these regions. In the commodity sector, crude oil prices reflected high sensitivity to geopolitical and market fundamentals, whereas gold and silver served as more stable stores of value with lower volatility. The most striking volatility was evident in the cryptocurrency market. Bitcoin and Ethereum registered exceptionally high standard deviations, underscoring the speculative fervor and rapid expansion phase of digital assets during this window. Conversely, major currency pairs maintained high levels of stability with minimal fluctuations. In summary, this period established a baseline of traditional market predictability against a backdrop of emerging digital asset turbulence.

**Table 1.** Descriptive statistics of the 12 products for the period between 1 January 2016 and 30 June 2018.

| | N | Average | Median | Std | Min | Max |
|---|---|---|---|---|---|---|
| S&P500 | 628 | 2,360.37 | 2,365.41 | 258.18 | 1,829.08 | 2,872.87 |
| DAX | 632 | 11,575.03 | 12,002.46 | 1,243.42 | 8,752.87 | 13,559.60 |
| Nikkei225 | 613 | 19,315.52 | 19,383.84 | 2,355.25 | 14,952.02 | 24,124.15 |
| Crude oil | 626 | 50.81 | 49.56 | 9.54 | 26.21 | 74.15 |
| Gold | 625 | 1,266.21 | 1,271.50 | 57.64 | 1,073.90 | 1,364.90 |
| Silver | 625 | 16.98 | 16.87 | 1.25 | 13.74 | 20.67 |
| Bitcoin | 912 | 3,649.33 | 1,187.47 | 4,189.45 | 364.33 | 19,497.40 |
| Ethereum | 911 | 233.20 | 44.89 | 298.89 | 0.92 | 1,398.99 |
| Litecoin | 912 | 52.79 | 6.93 | 72.31 | 3.00 | 358.34 |
| EUR/USD | 649 | 1.14 | 1.13 | 0.05 | 1.04 | 1.25 |

| | | | | | | |
|---|---|---|---|---|---|---|
| GBP/USD | 649 | 1.33 | 1.32 | 0.07 | 1.20 | 1.48 |
| AUD/USD | 649 | 0.76 | 0.76 | 0.02 | 0.69 | 0.81 |

Source: own editing.

During the initial observation window (Table 1), equity markets were characterized by steady expansion and generally contained price dispersion. While the S&P 500 index maintained an average level of 2,360.37 with a standard deviation of 258.18, both the DAX and Nikkei 225 registered higher price levels and greater variability, signaling more pronounced fluctuations in European and Asian markets. Crude oil prices averaged USD 50.81, though a standard deviation of USD 9.54 highlights the vulnerability of the energy sector to shifting geopolitical pressures and fundamental supply-demand dynamics. Conversely, precious metals like gold and silver acted as conservative hedges, demonstrating significantly lower volatility; silver's standard deviation, for instance, was restricted to USD 1.25.

The digital asset class, led by Bitcoin and Ethereum, stood out due to extreme volatility recorded at USD 4189.45 and USD 298.89, respectively reflecting the speculative intensity and explosive growth phases of cryptocurrencies during this period. In the foreign exchange markets, major pairs such as EUR/USD, GBP/USD, and AUD/USD remained remarkably resilient, trading within narrow corridors with minimal deviations.

**Table 2.** Descriptive statistics of the 12 products for the period between 1 January 2018 and 30 June 2020.

| | N | Average | Median | Std | Min | Max |
|---|---|---|---|---|---|---|
| S&P500 | 628 | 2,862.46 | 2,843.11 | 195.87 | 2,237.40 | 3,386.15 |
| DAX | 627 | 12,099.54 | 12,253.15 | 911.70 | 8,441.71 | 13,789.00 |
| Nikkei225 | 606 | 21,865.58 | 21,874.23 | 1,291.52 | 16,552.83 | 24,270.62 |
| Crude oil | 628 | 56.18 | 58.30 | 13.28 | -37.63 | 76.41 |
| Gold | 627 | 1,394.23 | 1,328.10 | 159.39 | 1,176.20 | 1,793.00 |
| Silver | 627 | 16.06 | 16.14 | 1.29 | 11.73 | 19.39 |
| Bitcoin | 912 | 7,679.92 | 7,679.97 | 2,384.93 | 3,236.76 | 17,527.00 |
| Ethereum | 911 | 302.16 | 207.80 | 238.30 | 81.72 | 1,398.99 |
| Litecoin | 912 | 80.24 | 61.06 | 48.48 | 23.46 | 296.45 |
| EUR/USD | 651 | 1.14 | 1.13 | 0.04 | 1.07 | 1.25 |
| GBP/USD | 651 | 1.30 | 1.29 | 0.05 | 1.15 | 1.43 |
| AUD/USD | 651 | 0.71 | 0.71 | 0.04 | 0.57 | 0.81 |

Source: authors' research.

The second observation window (Table 2), which encompasses the initial phase of the COVID-19 pandemic, was marked by rising volatility and a broad structural shift across global financial systems. While leading equity indices like the S&P 500 and DAX maintained their growth trajectories, a slight contraction in their standard deviations suggested a brief period of consolidation prior to the pandemic-induced turbulence. The most significant anomaly occurred in the energy sector, where crude oil futures plummeted to an unprecedented historic low of -

37.63 USD in April 2020. In response to this escalating global uncertainty, gold reinforced its status as a premier safe-haven asset, with its average price climbing to 1394.23 USD and its price dispersion (standard deviation) increasing substantially to 159.39 USD. Digital assets continued to exhibit high stochastic behavior, though rising median values suggested the early stages of market maturation. Meanwhile, currency markets reflected shifting economic vulnerabilities; while major pairs saw mild depreciation, the AUD/USD experienced a more significant downturn due to the Australian economy's heavy reliance on volatile commodity exports. Ultimately, this period was defined by extreme uncertainty and a dual-track process of adaptation within both conventional and emerging financial markets.

**Table 3.** Descriptive statistics of the 12 products for the period between 1 January 2020 and 30 June 2022.

| | N | Average | Median | Std | Min | Max |
|---|---|---|---|---|---|---|
| S&P500 | 628 | 3,851.17 | 3,915.98 | 599.47 | 2,237.40 | 4,796.56 |
| DAX | 635 | 13,907.57 | 13,950.04 | 1,633.03 | 8,441.71 | 16,271.75 |
| Nikkei225 | 606 | 26,031.83 | 27,007.45 | 3,214.96 | 16,552.83 | 30,670.10 |
| Crude oil | 628 | 63.11 | 62.57 | 25.23 | -37.63 | 123.70 |
| Gold | 628 | 1,803.00 | 1,809.30 | 104.27 | 1,477.30 | 2,051.50 |
| Silver | 628 | 22.99 | 23.97 | 3.70 | 11.73 | 29.40 |
| Bitcoin | 911 | 30,776.76 | 33,798.01 | 18,183.70 | 4,970.79 | 67,566.83 |
| Ethereum | 911 | 1,745.93 | 1,790.25 | 1,384.34 | 109.21 | 4,800.00 |
| Litecoin | 911 | 116.99 | 109.43 | 68.26 | 30.93 | 386.45 |
| EUR/USD | 651 | 1.15 | 1.16 | 0.05 | 1.04 | 1.23 |
| GBP/USD | 651 | 1.32 | 1.33 | 0.06 | 1.15 | 1.42 |
| AUD/USD | 651 | 0.72 | 0.72 | 0.04 | 0.57 | 0.80 |

Source: authors' research.

The third observation window (Table 3) was defined by a complex confluence of post-pandemic recovery and the geopolitical fallout from the Russian-Ukrainian conflict. Equity markets, while reaching higher average valuations, experienced a significant surge in stochastic behavior; for instance, the standard deviation of the S&P 500 escalated to 599.47 points, signaling a regime of heightened uncertainty. The energy sector proved particularly sensitive to the disruption of Russian exports, with crude oil reaching a peak of 123.70 USD.

In response to these systemic shocks, gold solidified its role as a safe-haven asset, maintaining an elevated average price of 1803.00 USD. Simultaneously, the cryptocurrency market hit record valuations, though Bitcoin's exceptionally high standard deviation of 18183.70 USD reflects a mixture of intense speculation and increased risk aversion in traditional markets. While major European currencies saw slight depreciation due to direct conflict proximity, the period was ultimately characterized by a transition toward high-volatility regimes and energy-driven market shocks.

The empirical analysis is structured across three partially overlapping sub-periods: 2016–2018, 2018–2020, and 2020–2022. This overlapping design was adopted for two primary reasons: (i)

Training Depth: It ensures that the deep learning architectures have access to a sufficiently longitudinal dataset for effective parameter estimation. (ii) Regime Continuity: It facilitates a more nuanced capture of market transitions, reflecting the reality that volatility shifts in financial time series are typically continuous rather than discrete.

To maintain the integrity of our results, each period and model is treated as a self-contained unit. Training and testing are performed exclusively within their respective windows, thereby eliminating the risk of information leakage across different regimes. This approach ensures that our findings regarding out-of-sample performance remain robust and independent.

### 3.2. Deep Learning Architectures

#### *3.2.1. Simple Recurrent Neural Network (RNN)*

A simple recurrent neural network (RNN) is a type of artificial neural network designed to capture temporal dependencies in sequential data. It consists of three main components: input, hidden, and output layers. Unlike traditional feedforward networks, RNNs incorporate recurrent connections within the hidden layer, allowing information from previous time steps to be retained and integrated into the current state. Specifically, the hidden state at time $t$ depends not only on the current input but also on the hidden state from the previous time step. This recursive structure enables the model to represent dynamic temporal behaviour in time series data, as past information can influence current predictions. As a result, RNNs are particularly well suited for modelling sequential patterns, where the ability to utilise previously learned information improves the efficiency and interpretability of the forecasting process [28].

The RNN can be formally defined as:

$$h_t = \phi(Ux_t + Wh_{t-1} + b_h) \tag{1}$$
$$y_t = f(Vh_t + b_y) \tag{2}$$

where $x_t \in \mathbb{R}^d$ denotes the input vector at time $t$, $h_t \in \mathbb{R}^m$ is the hidden state, and $y_t$ is the output. The matrices $U \in \mathbb{R}^{m\times d}$, $W \in \mathbb{R}^{m\times m}$, and $V \in \mathbb{R}^{k\times m}$ represent input, recurrent, and output weights, respectively. The vectors $b_h$ and $b_y$ are bias terms. The function $\phi(\cdot)$ typically denotes a nonlinear activation function such as $\tanh$, while $f(\cdot)$ is the output activation function.

#### *3.2.2. Gated Recurrent Unit (GRU)*

A Gated Recurrent Unit (GRU) is an extension of the standard RNN architecture that introduces gating mechanisms to regulate the flow of information through the network. Similar to RNNs, GRUs process sequential data by maintaining a hidden state that evolves over time. However, they address the limitations of traditional RNNs, particularly the vanishing gradient problem, through a more controlled memory update process. The GRU architecture includes two key components: the update gate and the reset gate. The update gate determines the extent to which the previous hidden state is retained, while the reset gate controls how much past information is incorporated into the current candidate state. This structure allows the model to balance between preserving long-term dependencies and adapting to new inputs. Compared to more complex architectures such as LSTM, GRUs have a simpler structure with fewer parameters, which can lead to improved computational efficiency while still maintaining strong predictive performance in time series applications [29].

The update gate $z_t$and reset gate $r_t$are defined as:

$$z_t = \sigma(W_z h_{t-1} + U_z x_t + b_z) \quad (3)$$
$$r_t = \sigma(W_r h_{t-1} + U_r x_t + b_r) \quad (4)$$

The candidate hidden state is given by:

$$\tilde{h}_t = \tanh(W_h(r_t \odot h_{t-1}) + U_h x_t + b_h) \quad (5)$$

The final hidden state is computed as:

$$h_t = z_t \odot h_{t-1} + (1 - z_t) \odot \tilde{h}_t \quad (6)$$

where $\sigma(\cdot)$denotes the sigmoid activation function, and $\odot$ represents element-wise multiplication. The gating structure allows the GRU to selectively retain or update information from previous states.

*3.2.3. Long-Short Term Memory (LSTM)*

The Long Short-Term Memory (LSTM) network is a specialised recurrent neural network architecture designed to capture long-term dependencies in sequential data. It extends the standard RNN by introducing a memory cell and a set of gating mechanisms that explicitly control the flow of information over time. An LSTM unit consists of three gates: the input gate, the forget gate, and the output gate. The input gate regulates the incorporation of new information into the memory cell, the forget gate determines which information should be discarded, and the output gate controls the information passed to the next time step. This gated structure enables the network to selectively retain or discard information, thereby mitigating the vanishing gradient problem commonly observed in standard RNNs. As a result, LSTM networks are particularly effective in modelling complex temporal patterns where long-term dependencies play a critical role, making them widely used in financial time series forecasting and other sequential learning tasks [30].

The input, forget, and output gates are defined as:

$$i_t = \sigma(W_{xi} x_t + W_{hi} h_{t-1} + b_i) \quad (7)$$
$$f_t = \sigma\left(W_{xf} x_t + W_{hf} h_{t-1} + b_f\right) \quad (8)$$
$$o_t = \sigma(W_{xo} x_t + W_{ho} h_{t-1} + b_o) \quad (9)$$

The candidate cell state is:

$$\tilde{c}_t = \tanh(W_{xc} x_t + W_{hc} h_{t-1} + b_c) \quad (10)$$

The cell state update is:

$$c_t = f_t \odot c_{t-1} + i_t \odot \tilde{c}_t \tag{11}$$

The hidden state is computed as:

$$h_t = o_t \odot \tanh(c_t) \tag{12}$$

where $c_t$represents the cell state, which serves as the memory component of the network.

*3.2.4.* Hybrid Architectures

To explore whether combining different recurrent structures improves forecasting performance, several hybrid architectures are implemented.

LSTM–GRU

The LSTM–GRU architecture combines the long-term memory capabilities of LSTM networks with the computational efficiency of GRU units. In this configuration, the input sequence is first processed by an LSTM layer, which captures long-term dependencies and produces a sequence of hidden representations. These representations are then passed to a GRU layer, which refines the extracted features through a more compact gating mechanism [32].

Formally, the model can be expressed as:

$$h_t^{(1)} = \mathrm{LSTM}\left(x_t, h_{t-1}^{(1)}\right) \tag{13}$$
$$h_t^{(2)} = \mathrm{GRU}\left(h_t^{(1)}, h_{t-1}^{(2)}\right) \tag{14}$$

The final prediction is obtained through a fully connected (dense) layer:

$$y_t = f(W_o h_t^{(2)} + b_o) \tag{15}$$

This architecture aims to balance expressive power and computational efficiency while reducing the risk of overfitting.

RNN–LSTM

The RNN–LSTM model combines a simple recurrent layer with an LSTM layer. The RNN component captures short-term temporal patterns, while the subsequent LSTM layer models longer-term dependencies based on the transformed input sequence [33].

$$h_t^{(1)} = \mathrm{RNN}\left(x_t, h_{t-1}^{(1)}\right) \tag{16}$$
$$h_t^{(2)} = \mathrm{LSTM}\left(h_t^{(1)}, h_{t-1}^{(2)}\right) \tag{17}$$
$$y_t = f\left(W_o h_t^{(2)} + b_o\right) \tag{18}$$

This structure allows the model to first extract basic temporal dynamics and then refine them using a more flexible memory mechanism.

RNN–GRU

The RNN–GRU architecture follows a similar design, where a standard RNN layer is followed by a GRU layer. This configuration provides a lightweight alternative to the RNN–LSTM model, with fewer parameters and lower computational cost.

$$h_t^{(1)} = \mathrm{RNN}\left(x_t, h_{t-1}^{(1)}\right) \tag{19}$$
$$h_t^{(2)} = \mathrm{GRU}\left(h_t^{(1)}, h_{t-1}^{(2)}\right) \tag{20}$$
$$y_t = f\left(W_o h_t^{(2)} + b_o\right) \tag{21}$$

Despite its simpler structure, the GRU layer enables the model to capture longer-term dependencies more effectively than a standard RNN alone.

These hybrid architectures are not intended to introduce fundamentally new model classes, but rather to assess whether combining different recurrent mechanisms improves robustness under varying market conditions.

### *3.3. Hyperparameters*

Table 4 presents the hyperparameter settings employed for the base models, which were also consistently applied to the hybrid architectures. To ensure comparability across model architectures, the study adopts a consistent set of hyperparameters based on commonly used values in the existing literature rather than performing model-specific optimisation. This approach allows for a controlled benchmarking environment in which differences in performance can be attributed primarily to architectural characteristics rather than to differences in tuning intensity. Limited experimentation is conducted with activation functions and optimisation settings in order to ensure stable convergence across models. While this design choice supports comparability, it also implies that the reported performance differences should be interpreted with appropriate caution. In particular, the results do not necessarily reflect the fully optimised potential of each architecture, but rather their relative behaviour under standardised conditions. This is especially relevant when comparing simpler and hybrid models, where dedicated hyperparameter tuning could potentially alter relative rankings.

**Table 4.** Hyperparameters.

| Model | Parameters | Value |
|---|---|---|
| RNN | Hidden Layers | 2 |
| | Hidden layer neuron count | 150 |
| | Batch size | 16 |
| | Epochs | 100 |
| | Activation | ReLU, Tanh, sigmoid |

| | | |
|---|---|---|
| | Learning rate | 0.001 |
| | Optimizer | Adam |
| LSTM | Hidden Layers | 2 |
| | Hidden layer neuron count | 150 |
| | Batch size | 16 |
| | Epochs | 100 |
| | Activation | ReLU, Tanh, sigmoid |
| | Learning rate | 0.001 |
| | Optimizer | Adam |
| GRU | Hidden Layers | 2 |
| | Hidden layer neuron count | 150 |
| | Batch size | 16 |
| | Epochs | 100 |
| | Activation | ReLU, Tanh, sigmoid |
| | Learning rate | 0.001 |
| | Optimizer | Adam |

Source: authors' research.

### *3.4. Software*

The computational framework for our model development was built using the Python programming language (version 3.9). Several technical and practical considerations justified this choice:

- Ecosystem and Libraries: Python provides a sophisticated and rapidly advancing ecosystem for data analytics and machine learning. Specifically, libraries such as Pandas and Numpy facilitate efficient data handling, while frameworks like TensorFlow, Keras, Pytorch, and Scikit-learn enable the implementation of cutting-edge neural network architectures.
- Accessibility and Reproducibility: As an open-source and platform-independent language, Python supports cost-effective research and ensures that experimental results can be consistently reproduced across different computing environments.
- Agile Development: The language's intuitive and readable syntax is highly conducive to rapid prototyping. This agility is vital in research settings where models require frequent iteration and precise parameter fine-tuning.
- Scalability: Python's extensive range of specialized frameworks, dedicated support for deep learning, and broad industrial applicability further reinforce its suitability for financial time series forecasting.

The development and execution of the scripts were conducted using a hybrid approach, utilizing the Google Colab interface for cloud-based tasks and Visual Studio Code for local development.

### *3.5. Forecasting Framework and Evaluation Methodology*

#### *3.5.1. Rolling Window Forecasting Design*

The forecasting procedure is implemented within a rolling-window framework designed to ensure realistic out-of-sample evaluation. Let $\{x_t\}_{t=1}^{T}$ denote the observed time series for a given asset. At each time step $t$, the model is trained using a fixed-length input sequence of $L = 50$ observations:

$$X_t = (x_{t-L+1}, x_{t-L+2}, \dots, x_t) \quad (22)$$

Based on this information set, the model produces a one-step-ahead forecast:

$$\hat{x}_{t+1} = f_\theta(X_t) \quad (23)$$

where $f_\theta(\cdot)$ denotes the nonlinear mapping learned by the neural network with parameters $\theta$. The window is then shifted forward by one time step, and the model is re-estimated iteratively using a walk-forward approach. This procedure ensures that all forecasts are generated strictly out-of-sample, as only information available at time $t$is used for predicting $t + 1$.

### *3.5.2. Univariate and Multivariate Specifications*

Two alternative input configurations are considered to assess the impact of information richness on forecasting performance. In the univariate setting, the input sequence consists solely of daily closing prices. Thus, each prediction is based on the past 50 observations of the closing price series. In the multivariate setting, the input space is extended to include additional market information. Specifically, opening, highest, and lowest prices are incorporated alongside closing prices, forming a multivariate input sequence over the same 50-day window. This design allows for a direct comparison between parsimonious and information-rich model specifications under identical forecasting conditions.

### *3.5.3. Train–Test Split, Model Re-estimation and Evaluation*

For each asset and period, the dataset is divided into training and test subsets using an 80–20 split:

$$T = T_{\text{train}} + T_{\text{test}},\ T_{\text{train}} = 0.8T,\ T_{\text{test}} = 0.2T \quad (24)$$

As a result, the evaluation periods are concentrated in the first halves of 2018, 2020, and 2022. Within this framework, model estimation follows a rolling (walk-forward) scheme. Let $t_0$ denote the starting point of the test period. At each time step $t \geq t_0$, the model is re-estimated using all information available up to that point. The corresponding training set is defined as:

$$D_t = \{x_1, x_2, \dots, x_t\} \quad (25)$$

Based on this information set, the model produces a one-step-ahead forecast:

$$\hat{x}_{t+1} = f_{\theta_t}(X_t) \quad (26)$$

where $\theta_t$ denotes the parameter set estimated at time $t$, and $X_t$ is the input sequence of length $L = 50$. This expanding-window re-estimation procedure ensures that forecasts are generated in a strictly out-of-sample manner, as only information available at time $t$is used for predicting $t + 1$. The approach reflects the evolving information set faced by market participants and eliminates the possibility of information leakage. To ensure comparability across assets and periods, all models are estimated independently for each asset-period combination. This design ensures that model performance reflects genuine predictive ability rather than artefacts arising from static sample splits or potential data leakage.

*3.5.4. Data Normalisation*

Prior to model estimation, all input variables are normalised using Min-Max scaling:

$$x_t^* = \frac{x_t - \min(X_{\text{train}})}{\max(X_{\text{train}}) - \min(X_{\text{train}})} \tag{27}$$

The scaling parameters are computed exclusively from the training data and applied consistently to the corresponding test observations, ensuring that no future information is incorporated into the transformation.

*3.5.5.* Forecast Accuracy Measure

Forecasting performance is evaluated using the Mean Absolute Percentage Error (MAPE):

$$\text{MAPE} = \frac{1}{N} \sum_{t=1}^{N} \left| \frac{y_t - \hat{y}_t}{y_t} \right| \tag{28}$$

where $y_t$and $\hat{y}_t$denote the actual and predicted values, respectively, and $N$ is the number of forecasts in the test period. For each asset and model, the reported MAPE value represents the average error across all rolling forecasts within the evaluation period.

*3.5.6. Cross-Sectional and Panel Regression Framework*

Cross-Sectional specification

To examine the relationship between market volatility and forecasting error, cross-sectional regressions are estimated for each period and model. Let $i = 1, \dots, N$ index the assets. The baseline cross-sectional regression model is specified as:

$$\text{MAPE}_i = \alpha + \beta \cdot \text{Volatility}_i + \varepsilon_i \tag{29}$$

where:
- $\text{MAPE}_i$denotes the average forecasting error for asset $i$,
- $\text{Volatility}_i$is the corresponding volatility measure,
- $\beta$captures the sensitivity of forecasting error to volatility.

These regressions are estimated separately for each model and period, allowing for a direct comparison of volatility sensitivity across architectures under different market conditions.

Panel Specification. In addition to the period-specific cross-sectional regressions, a pooled panel regression is estimated to capture the overall relationship across both assets and time periods.

Let $t = 1, \dots, T$ index the periods. The panel model is specified as:

$$\text{MAPE}_{i,t} = \alpha + \beta \cdot \text{Volatility}_{i,t} + \mu_i + \lambda_t + \varepsilon_{i,t} \tag{30}$$

where:

- $\mu_i$denotes asset-specific fixed effects,
- $\lambda_t$captures period-specific effects,
- $\varepsilon_{i,t}$is the idiosyncratic error term.

This specification allows controlling for unobserved heterogeneity across assets and common shocks across time periods.

*3.5.7. Sensitivity Analysis*

To assess the robustness of the estimated relationship, additional specifications are estimated by systematically excluding cryptocurrencies and commodities from the cross-sectional sample. The regression model in Equation (31) is re-estimated under these alternative samples:

$$\text{MAPE}_i^{(-g)} = \alpha^{(-g)} + \beta^{(-g)} \cdot \text{Volatility}_i^{(-g)} + \varepsilon_i \tag{31}$$

where $g$ denotes the excluded asset group. These specifications allow for evaluating whether the results are driven by specific high-volatility asset classes.

### *3.6. Volatility Analysis*

Because the raw variance across various financial products and timeframes is not conducive to direct comparison with MAPE values, a standardized indicator of relative volatility was required. To prevent price trends and periodic cycles from skewing the results, we first isolated and removed these components from the data. This objective was achieved through the application of the Seasonal-Trend Decomposition using Loess (STL) framework.

*3.6.1. STL (Seasonal-Trend decomposition using Loess)*

STL is a versatile time series decomposition technique that separates an observed series into its primary constituents: the long-term trend, the seasonal effect, and a random residual. A defining feature of this method is its reliance on Locally Weighted Scatterplot Smoothing (LOESS), which enables the model to capture complex, time-evolving, and nonlinear seasonal dynamics.

Its inherent robustness and adaptability make it particularly appropriate for the erratic nature of financial time series analysis.

The formal decomposition of the series is expressed as follows:

$$Y_t = T_t + S_t + R_t \tag{32}$$

In this expression, $Y_t$ signifies the raw observation series, while $T_t$, $S_t$ and $R_t$ represent the secular trend, the seasonal effect, and the random residual, respectively. The application of STL is highly effective for financial instruments characterized by non-stationary seasonal structures that are prone to temporal shifts [34]. Furthermore, the residual component $R_t$ encapsulates stochastic, short-term fluctuations that remain after the trend and seasonality have been filtered out, providing a robust proxy for targeted volatility analysis [35]

Direct comparison of variance across heterogeneous assets and periods is not appropriate for assessing volatility and forecasting error; consequently, a relative measure of volatility is introduced. The construction of this indicator is described in detail below. Prior to its computation, it is essential to isolate trend components in the time series, as their presence may distort the analysis. For this purpose, the Seasonal-Trend Decomposition using Loess (STL) method is applied, as discussed in the following sections.

### *3.6.2. Relative Volatility: Coefficient of Variation (CV)*

We quantified the volatility of the trend-filtered series using the Coefficient of Variation (CV). As a dimensionless metric, the CV enables a consistent comparison of dispersion across assets characterized by vastly different scales, units of measurement, and price magnitudes.

The calculation is defined by the following ratio:

$$CV = \frac{\sigma(R_t)}{\mu(Y_t)} \tag{33}$$

Where: $(R_t)$ is the standard deviation of the residuals derived from the STL decomposition and $\mu(Y_t)$ denotes the arithmetic mean of the original price series.

While the CV is sensitive to the magnitude of the series average potentially indicating higher relative volatility in assets with lower nominal values it provides a balanced bridge between absolute and relative variance measures [36]. Furthermore, this indicator serves as a robust foundation for subsequent cluster analysis. Ultimately, this metric provides a standardized proxy for volatility derived from residuals that have been successfully stripped of seasonal and trend-related distortions.

## 4. Results

The investigation into the nexus between market volatility and forecasting accuracy was initiated following the systematic optimization of activation functions. An analysis of relative volatility metrics reveals that digital assets exhibited the most pronounced fluctuations, particularly during 2018, with Ethereum and Litecoin recording coefficients of 0.0475 and

0.0432, respectively. These figures stand in sharp contrast to the stability observed in foreign exchange markets, exemplified by the EUR/USD pair's low CV of 0.0032. The 2020 pandemic-induced market shock resulted in an unprecedented volatility spike in crude oil, which reached 0.1005. Furthermore, major equity benchmarks, including the S&P 500 and DAX, demonstrated elevated volatility levels during the 2020 crisis relative to the 2018 or 2022 windows, reflecting a significant intensification of systemic uncertainty.

Pooled regression analysis across the entire observation timeframe confirms a significant positive correlation between relative volatility (CV_STL) and univariate forecasting error (MAPE). The estimated regression coefficient of 1.2937 ($p<0.01$) indicates that a single-unit rise in relative volatility corresponds to an average increase of approximately 1.29 units in prediction error. While the constant term was recorded at 0.0049, its lack of statistical significance suggests that market volatility is the primary determinant of MAPE variability. The model demonstrated robust explanatory power, yielding an $R^2$ value of 76.28%. Regarding specific error rates in 2018, univariate models performed least effectively for Litecoin (0.1095) and Bitcoin (0.1011). Conversely, the highest precision was achieved in the currency markets, with GBP/USD and AUD/USD recording MAPE values as low as 0.0047 and 0.0055, respectively. The disruptive impact of volatility was further underscored by the 2020 crude oil data, which yielded a high MAPE of 0.1023, confirming that extreme price movements fundamentally degrade forecasting reliability. Detailed asset- and model-specific univariate MAPE results for the 2018, 2020, and 2022 evaluation periods are presented in Appendix B.

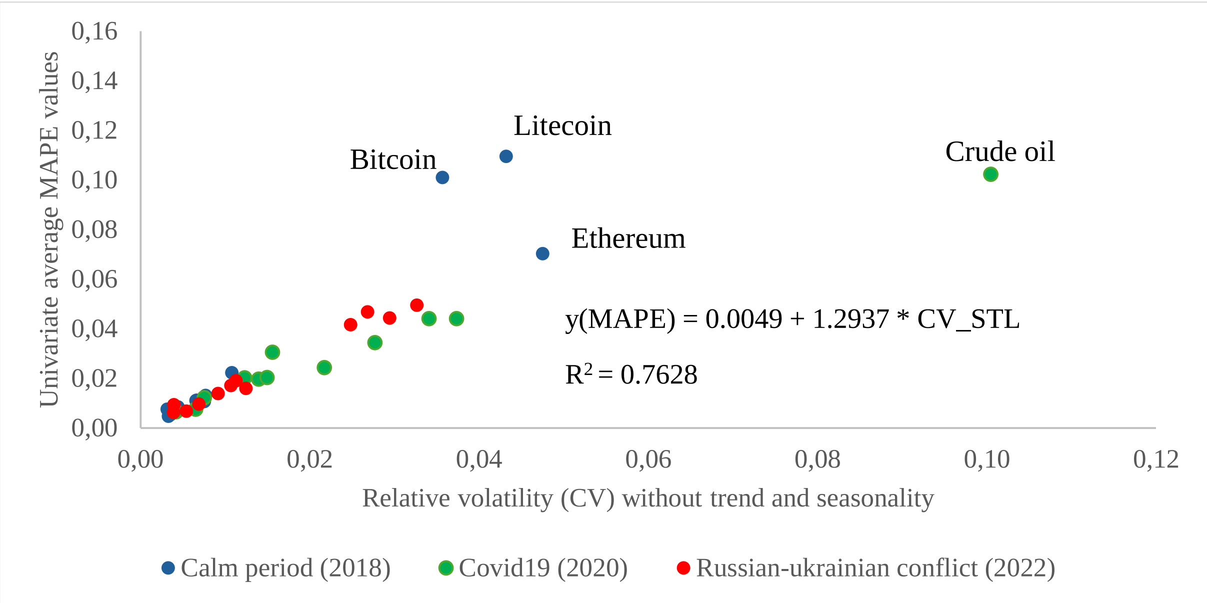


**Figure 1.** The relationship between relative volatility and univariate average MAPE values, displayed by time periods.
Source: authors' research.

The influence of market turbulence is even more striking within the multivariate specification (Figure 2). In this model, the regression coefficient for CV_STL escalates to 1.6224 ($p<0.01$), indicating that multivariate forecasting systems exhibit a heightened sensitivity to volatility compared to their univariate counterparts. This finding suggests that increasing the complexity of the input layer does not inherently provide a buffer against market shocks; instead, it may amplify error sensitivity during periods of rapid price reconfiguration. Similar to the univariate results, the constant term (-0.0009) remains statistically non-significant, further reinforcing the conclusion that relative volatility serves as the decisive determinant of model variability. The explanatory power of this framework is notably superior to the univariate baseline, yielding an $R^2$ of 90.58%. While the multivariate models generally produce lower or comparable error rates across the sample, specific cases reveal the risks of overparameterization. For instance, in 2018, the multivariate MAPE for Litecoin rose to 0.1198, exceeding its univariate error. This anomaly likely stems from challenges in predictor selection or the breakdown of intraday correlation structures. Despite these outliers, the multivariate approach provided incremental gains in accuracy for several instruments, such as the Nikkei225 in 2020 (0.0203 vs. 0.0204) and the GBP/USD exchange rate in 2022 (0.0040 vs. 0.0063). The corresponding asset- and model-specific multivariate MAPE results are reported in Appendix B.

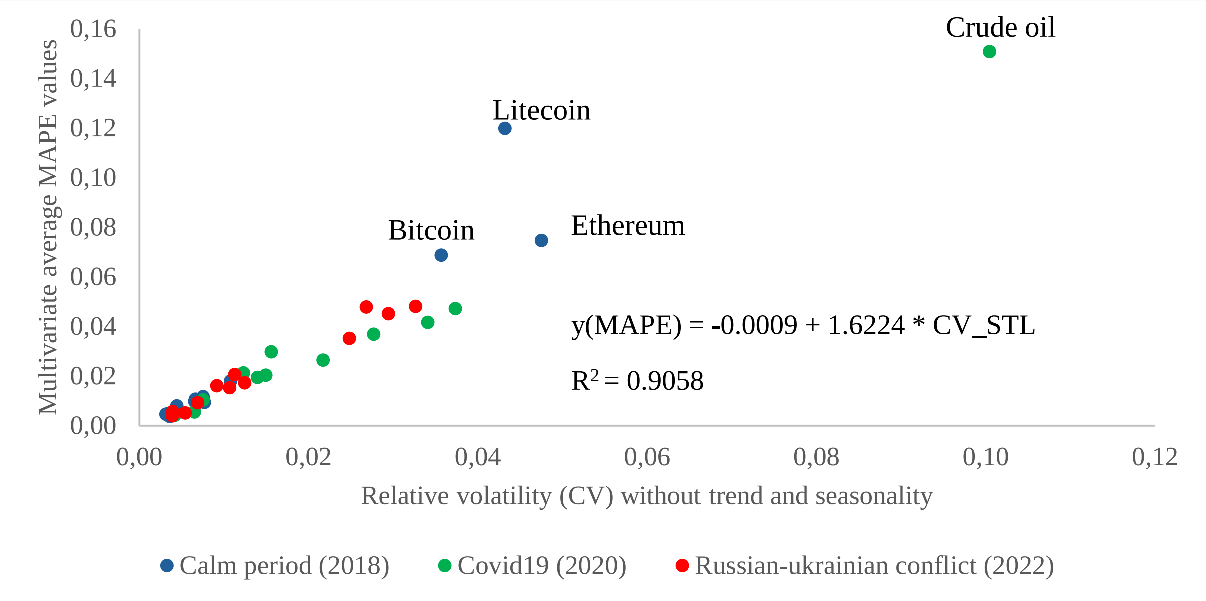


**Figure 2.** The relationship between relative volatility and multivariate average MAPE values, displayed by time periods.
Source: authors' research.

The granular relationship between relative volatility and MAPE, categorized by model architecture across the 2018, 2020, and 2022 periods, is presented in Tables 4–6. Table 5 specifically details the interaction between CV_STL and forecasting errors for six neural network configurations during the 2018 stable regime.

Our findings confirm a universally significant and positive correlation between volatility levels and univariate MAPE across all tested architectures. This relationship dictates that as the inherent volatility of a financial asset escalates, the associated prediction error increases commensurately. However, the degree of this sensitivity varies markedly by model type:

- LSTM models exhibited the highest sensitivity to volatility, with a regression coefficient of 2.920.
- GRU models proved significantly more robust, recording the lowest coefficient of 1.528.
- In terms of goodness-of-fit, the GRU architecture achieved the most substantial explanatory power with an $R^2$ of 0.9713, while the RNN baseline showed the weakest fit at 0.6134.

These results substantiate the role of relative volatility as a consistent and powerful predictor of univariate forecasting performance. The transition to a multivariate framework does not decouple the fundamental link between market turbulence and forecasting inaccuracy. Even with the inclusion of expanded input variables, more volatile assets continue to yield significantly higher prediction errors. The architectural hierarchy of sensitivity remained distinct in the multivariate setting:

- The RNN-LSTM hybrid model demonstrated the highest volatility sensitivity, yielding a regression coefficient of 2.856.
- The LSTM-GRU configuration emerged as the most resilient multivariate architecture, with a minimum coefficient of 1.662.
- Reflecting this resilience, the LSTM-GRU model provided the best statistical fit ($R^2$=0.9651), whereas the RNN-LSTM model exhibited the lowest explanatory power at 0.7951.

Ultimately, the 2018 data confirms that both univariate and multivariate architectures maintain a stable, positive relationship between relative volatility and prediction error, though simpler gating mechanisms often provide superior robustness.

**Table 5.** Regression results of the relative volatility indicators and MAPE values for the examined assets during the calm period (2018).

| | **(1)** | **(2)** | **(3)** | **(4)** | **(5)** | **(6)** |
|---|---|---|---|---|---|---|
| | **Univ-RNN** | **Univ-LSTM** | **Univ-GRU** | **Univ-LSTM-GRU** | **Univ-RNN-LSTM** | **Univ-RNN-GRU** |
| CV_STL | $2.422^{***}$ | $2.920^{***}$ | $1.528^{***}$ | $2.128^{***}$ | $2.304^{***}$ | $1.817^{***}$ |
| | (0.608) | (0.578) | (0.083) | (0.315) | (0.168) | (0.275) |
| Constant | -0.002 | -0.006 | -0.000 | -0.001 | -0.003 | 0.002 |
| | (0.013) | (0.013) | (0.002) | (0.007) | (0.004) | (0.006) |
| Observations | 12 | 12 | 12 | 12 | 12 | 12 |
| $R^2$ | 0.6134 | 0.7185 | 0.9713 | 0.8207 | 0.9498 | 0.8132 |
| | **(7)** | **(8)** | **(9)** | **(10)** | **(11)** | **(12)** |
| | **Multi-RNN** | **Multi-LSTM** | **Multi-GRU** | **Multi-LSTM-GRU** | **Multi-RNN-LSTM** | **Multi-RNN-GRU** |
| CV_STL | $1.830^{***}$ | $2.854^{***}$ | $1.810^{***}$ | $1.662^{***}$ | $2.856^{***}$ | $1.870^{***}$ |
| | (0.264) | (0.346) | (0.188) | (0.100) | (0.458) | (0.239) |

| | | | | | | |
|---|---|---|---|---|---|---|
| Constant | 0.000 | -0.008 | -0.004 | -0.001 | -0.007 | -0.001 |
| | (0.006) | (0.008) | (0.004) | (0.002) | (0.010) | (0.005) |
| Observations | 12 | 12 | 12 | 12 | 12 | 12 |
| $R^2$ | 0.8273 | 0.8720 | 0.9022 | 0.9651 | 0.7951 | 0.8595 |

Note: In parentheses, the standard errors $^{*}$ $p < 0.1$, $^{**}$ $p < 0.05$, $^{***}$ $p < 0.01$.
Source: authors' research based on STATA 17 results.

The interaction between relative volatility (CV_STL) and forecasting errors during the 2020 pandemic shock is detailed in Table 6, covering six distinct neural network configurations. Within the univariate framework, every model yielded a significant and positive regression coefficient for volatility, affirming that heightened market turbulence consistently leads to increased prediction errors during crisis regimes. Architectural differences were notable, as the GRU model demonstrated the lowest volatility sensitivity with a coefficient of 0.845, while the RNN baseline exhibited the highest at 1.158. The explanatory power of these regressions was exceptionally robust, with $R^2$ values exceeding 90% for all architectures, indicating that volatility served as a dominant factor in determining error levels during this turbulent period.

The multivariate results further underscore this relationship, with all specifications maintaining a statistically significant and positive link between CV_STL and MAPE. In this more complex input setting, the GRU architecture again displayed superior robustness, recording the lowest sensitivity at 0.828, whereas the RNN-LSTM hybrid proved the most vulnerable with a coefficient of 2.847. The coefficients of determination remained remarkably high across the models, spanning a range of 86% to 98%, with the LSTM model specifically achieving the strongest fit at 0.9714. These empirical findings provide compelling evidence that even amidst the unique structural breaks of the 2020 pandemic, the fundamental impact of market fluctuations on forecasting performance persists across both standard and hybrid deep learning architectures.

**Table 6.** Regression results of the relative volatility indicators and MAPE values for the examined assets during the COVID-19 period (2020).

| | **(1) Univ-RNN** | **(2) Univ-LSTM** | **(3) Univ-GRU** | **(4) Univ-LSTM-GRU** | **(5) Univ-RNN-LSTM** | **(6) Univ-RNN-GRU** |
|---|---|---|---|---|---|---|
| CV_STL | $1.158^{***}$ | $0.964^{***}$ | $0.845^{***}$ | $0.984^{***}$ | $0.984^{***}$ | $0.936^{***}$ |
| | (0.067) | (0.044) | (0.077) | (0.046) | (0.100) | (0.064) |
| Constant | $0.005^{*}$ | $0.004^{**}$ | $0.006^{**}$ | $0.005^{***}$ | $0.011^{**}$ | $0.007^{***}$ |
| | (0.002) | (0.002) | (0.003) | (0.002) | (0.004) | (0.002) |
| Observations | 12 | 12 | 12 | 12 | 12 | 12 |
| $R^2$ | 0.9676 | 0.9795 | 0.9228 | 0.9787 | 0.9067 | 0.9549 |
| | **(7) Multi-RNN** | **(8) Multi-LSTM** | **(9) Multi-GRU** | **(10) Multi-LSTM-GRU** | **(11) Multi-RNN-LSTM** | **(12) Multi-RNN-GRU** |
| CV_STL | $2.279^{***}$ | $0.896^{***}$ | $0.828^{***}$ | $0.911^{***}$ | $2.847^{***}$ | $1.150^{***}$ |

| | | | | | | |
|---|---|---|---|---|---|---|
| | (0.158) | (0.049) | (0.066) | (0.050) | (0.289) | (0.146) |
| Constant | -0.010* | 0.004** | 0.005* | 0.004** | -0.020* | 0.005 |
| | (0.006) | (0.002) | (0.002) | (0.002) | (0.010) | (0.005) |
| Observations | 12 | 12 | 12 | 12 | 12 | 12 |
| $R^2$ | 0.9543 | 0.9714 | 0.9403 | 0.9706 | 0.9065 | 0.8616 |

Note: In parentheses, the standard errors * $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$.
Source: authors' research based on STATA 17 results.

Table 7 reports the association between relative volatility (CV_STL) and forecasting errors (MAPE) for both univariate and multivariate model specifications in 2022. In the case of univariate models, the estimated coefficients are uniformly positive and statistically significant, indicating that higher levels of volatility are systematically linked to increased prediction errors. The magnitude of the coefficients ranges from 1.149 for the GRU model to 1.969 for the RNN–LSTM architecture, suggesting a broadly consistent, moderate sensitivity to volatility across model types. The corresponding $R^2$ values are also high, falling between 0.85 and 0.92, which indicates substantial explanatory power. A similar pattern emerges for the multivariate specifications. All models produce positive and statistically significant coefficients, reinforcing the presence of a stable relationship between volatility and forecasting error. The estimated coefficients lie within a relatively narrow interval, with the GRU model exhibiting the lowest sensitivity (1.312) and the RNN–LSTM model the highest (2.033), reflecting heterogeneity in how different architectures respond to volatility. The explanatory power remains strong, with $R^2$ values ranging from 0.82 to 0.97, and the LSTM–GRU model achieving the highest fit (0.9748). Overall, the findings for 2022 are consistent with earlier periods, supporting the interpretation that volatility exerts a persistent and structurally embedded influence on forecasting performance, even when more complex multivariate inputs are employed.

**Table 7.** Regression results of the relative volatility indicators and univariate MAPE values for the examined assets during the Russian–Ukrainian conflict period (2022).

| | **(1) Univ-RNN** | **(2) Univ-LSTM** | **(3) Univ-GRU** | **(4) Univ-LSTM-GRU** | **(5) Univ-RNN-LSTM** | **(6) Univ-RNN-GRU** |
|---|---|---|---|---|---|---|
| CV_STL | 1.856*** | 1.403*** | 1.149*** | 1.539*** | 1.969*** | 1.598*** |
| | (0.205) | (0.132) | (0.133) | (0.192) | (0.254) | (0.165) |
| Constant | -0.002 | -0.000 | 0.003 | 0.000 | -0.001 | -0.001 |
| | (0.004) | (0.002) | (0.002) | (0.003) | (0.005) | (0.003) |
| Observations | 12 | 12 | 12 | 12 | 12 | 12 |
| $R^2$ | 0.8915 | 0.9192 | 0.8819 | 0.8656 | 0.8570 | 0.9032 |

| | **(7) Multi-RNN** | **(8) Multi-LSTM** | **(9) Multi-GRU** | **(10) Multi-LSTM-GRU** | **(11) Multi-RNN-LSTM** | **(12) Multi-RNN-GRU** |
|---|---|---|---|---|---|---|
| CV_STL | 1.370*** | 1.523*** | 1.312*** | 1.395*** | 2.033*** | 1.903*** |
| | (0.153) | (0.225) | (0.115) | (0.071) | (0.225) | (0.188) |

| | | | | | | |
|---|---|---|---|---|---|---|
| Constant | 0.003<br>(0.003) | -0.000<br>(0.004) | -0.001<br>(0.002) | -0.001<br>(0.001) | -0.003<br>(0.004) | -0.003<br>(0.003) |
| Observations | 12 | 12 | 12 | 12 | 12 | 12 |
| $R^2$ | 0.8889 | 0.8206 | 0.9283 | 0.9748 | 0.8904 | 0.9112 |

Note: In parentheses, the standard errors $^{*}$ $p < 0.1$, $^{**}$ $p < 0.05$, $^{***}$ $p < 0.01$.
Source: authors' research based on STATA 17 results.

Table 8 presents sensitivity analyses of the volatility-error relationship across alternative asset samples, based on robustness checks of the cross-sectional estimates. These additional specifications assess whether the observed relationships are driven by a small subset of highly volatile assets, such as cryptocurrencies or commodities, and to evaluate the stability of the estimated coefficients under alternative sample compositions. The results of the sensitivity analysis clearly indicate that the relationship between volatility and forecasting error cannot be attributed solely to specific asset classes, particularly not to cryptocurrencies or commodity assets. Across most models and periods, the sign of the volatility coefficient remains unchanged, and its magnitude is relatively stable even when high-volatility assets are excluded from the sample. This pattern is especially evident in GRU-based models, where the coefficients exhibit only minor variation while the direction and statistical significance of the relationship are preserved. Excluding cryptocurrencies reduces the magnitude of the coefficients in some cases (particularly in 2018), suggesting that these assets do contribute to amplifying the volatility effect. However, the relationship remains consistently detectable even after their removal. In contrast, excluding commodity assets generally results in smaller changes, indicating that these assets play a less dominant role in driving the results. Importantly, the $R^2$ values show only moderate variation across specifications, pointing to the relative stability of the models' explanatory power under different sample compositions. In the 2020 period, which represents a highly volatile market environment, the coefficients are particularly consistent across all three sample specifications, reinforcing the conclusion that the observed relationship is not driven solely by extreme asset classes. The results for 2022 exhibit a similar pattern, although somewhat larger variations are observed for certain models, potentially reflecting changes in market structure. These findings suggest that the positive relationship between volatility and forecasting error is structural in nature and not driven by a single subset of assets. These observations support the conclusion that the earlier findings are not sensitive to a narrow set of assets, thereby strengthening the generalisability of the results derived from the cross-sectional regressions beyond what would be implied by the full-sample estimates alone.

**Table 8.** Sensitivity Analysis of the Volatility-Error Relationship Across Alternative Asset Samples for all periods.

| Period | Model | Coefficients (β) full sample | $R^2$ | Coefficients (β) without cryptocurrencies | $R^2$ | Coefficients (β) without commodities | $R^2$ |
|---|---|---|---|---|---|---|---|
| 2018 | Univ-RNN | 2.422** (0.892) | 0.6134 | 1.110** (0.441) | 0.4793 | 2.369** (0.899) | 0.5862 |
| | Univ-LSTM | 2.920** (1.089) | 0.7185 | 1.737*** (0.234) | 0.7627 | 2.893** (1.123) | 0.7000 |
| | Univ-GRU | 1.528*** (0.124) | 0.9713 | 1.582*** (0.129) | 0.7915 | 1.530*** (0.127) | 0.9690 |
| | Univ-LSTM-GRU | 2.128*** (0.537) | 0.8207 | 1.832** (0.526) | 0.7164 | 2.125*** (0.551) | 0.8083 |
| | Univ-RNN-LSTM | 2.304*** (0.247) | 0.9498 | 2.000** (0.716) | 0.5966 | 2.327*** (0.250) | 0.9529 |

| | | | | | | | |
|---|---|---|---|---|---|---|---|
| | Univ-RNN-GRU | 1.817*** (0.432) | 0.8132 | 3.391*** (0.568) | 0.7828 | 1.799*** (0.439) | 0.8084 |
| | Multi-RNN | 1.830*** (0.433) | 0.8273 | 2.143*** (0.464) | 0.7986 | 1.848*** (0.443) | 0.8202 |
| | Multi-LSTM | 2.854*** (0.611) | 0.8720 | 2.273*** (0.371) | 0.8326 | 2.866*** (0.627) | 0.8644 |
| | Multi-GRU | 1.810*** (0.319) | 0.9022 | 1.434*** (0.114) | 0.9269 | 1.817*** (0.325) | 0.8968 |
| | Multi-LSTM-GRU | 1.662*** (0.154) | 0.9651 | 2.047*** (0.188) | 0.8531 | 1.679*** (0.157) | 0.9652 |
| | Multi-RNN-LSTM | 2.856*** (0.808) | 0.7951 | 0.981** (0.282) | 0.6449 | 2.828** (0.829) | 0.7816 |
| | Multi-RNN-GRU | 1.870*** (0.445) | 0.8595 | 1.238*** (0.136) | 0.8172 | 1.858*** (0.459) | 0.8490 |
| | Univ-RNN | 1.158*** (0.042) | 0.9676 | 1.137*** (0.015) | 0.9809 | 1.424*** (0.139) | 0.9512 |
| 2020 | Univ-LSTM | 0.964*** (0.023) | 0.9795 | 0.977*** (0.019) | 0.9843 | 0.906*** (0.110) | 0.8745 |
| | Univ-GRU | 0.845*** (0.081) | 0.9228 | 0.802*** (0.018) | 0.9762 | 1.285*** (0.128) | 0.9468 |
| | Univ-LSTM-GRU | 0.984*** (0.031) | 0.9787 | 0.974*** (0.019) | 0.9807 | 1.116*** (0.068) | 0.9268 |
| | Univ-RNN-LSTM | 0.984*** (0.055) | 0.9067 | 0.985*** (0.048) | 0.9045 | 1.194*** (0.126) | 0.8666 |
| | Univ-RNN-GRU | 0.936*** (0.035) | 0.9549 | 0.933*** (0.025) | 0.9668 | 1.021*** (0.180) | 0.8202 |
| | Multi-RNN | 2.279*** (0.182) | 0.9543 | 2.375*** (0.032) | 0.9917 | 1.267*** (0.189) | 0.8399 |
| | Multi-LSTM | 0.896*** (0.037) | 0.9714 | 0.891*** (0.031) | 0.9723 | 1.044*** (0.103) | 0.8997 |
| | Multi-GRU | 0.828*** (0.075) | 0.9403 | 0.794*** (0.026) | 0.9742 | 1.232*** (0.077) | 0.9599 |
| | Multi-LSTM-GRU | 0.911*** (0.030) | 0.9706 | 0.914*** (0.026) | 0.9764 | 1.005*** (0.140) | 0.8745 |
| | Multi-RNN-LSTM | 2.847*** (0.334) | 0.9065 | 3.017*** (0.066) | 0.9775 | 1.131*** (0.181) | 0.7919 |
| | Multi-RNN-GRU | 1.150*** (0.139) | 0.8616 | 1.077*** (0.019) | 0.9909 | 1.885*** (0.466) | 0.8433 |
| 2022 | Univ-RNN | 1.856*** (0.239) | 0.8915 | 2.388*** (0.228) | 0.9414 | 1.655*** (0.240) | 0.9240 |
| | Univ-LSTM | 1.403*** (0.178) | 0.9192 | 1.675*** (0.088) | 0.9818 | 1.316*** (0.183) | 0.9184 |
| | Univ-GRU | 1.149*** (0.166) | 0.8819 | 0.878*** (0.136) | 0.7505 | 1.244*** (0.158) | 0.9147 |
| | Univ-LSTM-GRU | 1.539*** (0.246) | 0.8656 | 1.898*** (0.160) | 0.9348 | 1.410*** (0.255) | 0.8565 |
| | Univ-RNN-LSTM | 1.969*** (0.355) | 0.8570 | 2.331*** (0.306) | 0.9132 | 1.796*** (0.380) | 0.8485 |
| | Univ-RNN-GRU | 1.598*** (0.235) | 0.9032 | 1.200*** (0.088) | 0.8801 | 1.684*** (0.242) | 0.9230 |
| | Multi-RNN | 1.370*** (0.112) | 0.8889 | 1.241*** (0.171) | 0.6887 | 1.400*** (0.114) | 0.8958 |
| | Multi-LSTM | 1.523*** (0.279) | 0.8206 | 2.430*** (0.147) | 0.9586 | 1.263*** (0.132) | 0.9244 |
| | Multi-GRU | 1.312*** (0.145) | 0.9283 | 1.066*** (0.143) | 0.8814 | 1.403*** (0.154) | 0.9549 |
| | Multi-LSTM-GRU | 1.395*** (0.077) | 0.9748 | 1.322*** (0.049) | 0.9754 | 1.432*** (0.086) | 0.9774 |
| | Multi-RNN-LSTM | 2.033*** (0.276) | 0.8904 | 2.891*** (0.259) | 0.9530 | 1.789*** (0.112) | 0.9812 |
| | Multi-RNN-GRU | 1.903*** (0.158) | 0.9112 | 2.411*** (0.111) | 0.8977 | 1.751*** (0.105) | 0.9238 |

Note: In parentheses, the robust standard errors $^{*}$ $p < 0.1$, $^{**}$ $p < 0.05$, $^{***}$ $p < 0.01$.
Source: authors' research based on STATA 17 results.

To facilitate a robust quantitative comparison across the various architectures, we performed a panel regression analysis integrating data from all observation windows for both univariate and multivariate specifications. The synthesized findings are documented in Table 9. When examining the relationship between univariate MAPE and relative volatility (CV_STL), our results consistently yielded positive and statistically significant coefficients at the $p<0.01$ level. This confirms that increased market volatility is systematically linked to higher forecasting errors within the panel framework. In terms of architectural sensitivity, the RNN model displayed the highest volatility response with a coefficient of 1.485, whereas the GRU model remained the most resilient with a value of 0.923. The goodness-of-fit metrics revealed considerable heterogeneity; the GRU architecture provided the most robust statistical fit ($R^2$=0.8632), while the LSTM model recorded the lowest explanatory power at 0.5259.

A parallel trend emerged in the multivariate configurations, where all aggregated panel regressions showed a significant positive association between relative volatility and error rates. Within this information-rich environment, the RNN-LSTM hybrid recorded the highest sensitivity to market turbulence with a coefficient of 2.719. Conversely, the GRU once again demonstrated superior robustness, yielding the lowest coefficient of 1.035. The explanatory power across these multivariate models was generally high, ranging from 60% to 90%, with the multivariate RNN achieving a particularly strong fit of 90.72%. Collectively, these panel-level insights underscore that volatility remains the primary driver of prediction errors for both univariate and multivariate deep learning systems, regardless of the time period or model complexity. This reinforces the thesis that market stability is the fundamental constraint on the predictive capabilities of even the most advanced algorithmic frameworks.

**Table 9.** Panel regression results of the relative volatility indicators and MAPE values for the examined assets across all periods.

| | **(1)** | **(2)** | **(3)** | **(4)** | **(5)** | **(6)** |
|---|---|---|---|---|---|---|
| | **Univ-RNN** | **Univ-LSTM** | **Univ-GRU** | **Univ-LSTM-GRU** | **Univ-RNN-LSTM** | **Univ-RNN-GRU** |
| CV_STL | $1.485^{***}$ | $1.418^{***}$ | $0.923^{***}$ | $1.274^{***}$ | $1.373^{***}$ | $1.185^{***}$ |
| | (0.201) | (0.231) | (0.067) | (0.135) | (0.131) | (0.114) |
| Constant | 0.004 | 0.003 | $0.006^{***}$ | 0.005 | $0.007^{*}$ | $0.006^{**}$ |
| | (0.005) | (0.006) | (0.002) | (0.004) | (0.003) | (0.003) |
| Observations | 36 | 36 | 36 | 36 | 36 | 36 |
| $R^2$ | 0.6157 | 0.5259 | 0.8632 | 0.7251 | 0.7645 | 0.7606 |
| | **(7)** | **(8)** | **(9)** | **(10)** | **(11)** | **(12)** |
| | **Multi-RNN** | **Multi-LSTM** | **Multi-GRU** | **Multi-LSTM-GRU** | **Multi-RNN-LSTM** | **Multi-RNN-GRU** |
| CV_STL | $2.144^{***}$ | $1.369^{***}$ | $1.035^{***}$ | $1.117^{***}$ | $2.719^{***}$ | $1.198^{***}$ |
| | (0.118) | (0.192) | (0.104) | (0.073) | (0.188) | (0.112) |
| Constant | $-0.007^{**}$ | 0.003 | 0.003 | 0.003 | $-0.012^{**}$ | $0.006^{*}$ |
| | (0.003) | (0.005) | (0.003) | (0.002) | (0.005) | (0.004) |
| Observations | 36 | 36 | 36 | 36 | 36 | 36 |
| $R^2$ | 0.9072 | 0.5982 | 0.7794 | 0.8734 | 0.8604 | 0.8102 |

Note: In parentheses, the standard errors $^{*}$ $p < 0.1$, $^{**}$ $p < 0.05$, $^{***}$ $p < 0.01$.
Source: authors' research based on STATA 17 results.

The consolidated panel regression results presented in Table 9 facilitate a comparative assessment of model sensitivity relative to fluctuations in CV_STL. The universally positive nature of these coefficients across all architectures reinforces the premise that intensifying market volatility directly correlates with elevated forecasting errors, as measured by MAPE. Consequently, the magnitude of these parameters serves as a critical proxy for structural robustness, where lower coefficients denote reduced sensitivity to stochastic market shifts.

A comparative analysis of input configurations reveals that multivariate frameworks typically exhibit higher sensitivity to volatility shocks than their univariate counterparts. Specifically, for the RNN, GRU, RNN-LSTM, and RNN-GRU architectures, the univariate specifications yielded more favorable, lower regression coefficients. In contrast, the LSTM and LSTM-GRU models followed an opposite pattern, showing improved stability in the multivariate setting. This creates a notable trade-off: although multivariate models often achieve superior predictive fit – evidenced by a Multi-RNN $R^2$ of 0.9072 compared to 0.6157 for the univariate version – they remain more susceptible to performance degradation during turbulent regimes. This vulnerability suggests that explicit volatility management is particularly warranted when deploying these more complex, information-rich models.

Examining the hierarchy between standalone and hybrid architectures provides further nuance. Within univariate specifications, hybrid versions of the RNN and LSTM displayed enhanced stability compared to their base counterparts. However, the GRU architecture behaved differently, with the standard base model exhibiting lower volatility sensitivity than its hybrid variants. This superior performance of the standalone GRU was also evident in the multivariate analysis. While the findings for RNN and LSTM hybrids in the multivariate context were more heterogeneous – with certain configurations like RNN-GRU and LSTM-GRU demonstrating gains in robustness – univariate models remain, on average, less reactive to volatility and thus more structurally stable.

A simpler architecture, and the GRU in particular, demonstrate lower volatility sensitivity compared to many hybrid configurations. The performance spectrum for hybrid models is broad; while certain architectures like the RNN-LSTM are highly susceptible to market noise, others such as the RNN-GRU exhibit surprising stability. These results emphasize that the selection of both the model architecture and the underlying input structure is paramount for maintaining reliability in highly volatile environments.

## 5. Discussion and Limitations

Robustness is interpreted in this study in a restricted sense, referring to the stability of the relationship between volatility and forecasting error across different asset samples and model specifications. This interpretation focuses on cross-sectional consistency rather than on performance stability under alternative resampling schemes or fully independent out-of-sample protocols. Accordingly, the results should not be interpreted as a comprehensive robustness assessment in the strict methodological sense, but rather as evidence on the persistence of the volatility–error relationship across heterogeneous financial assets.

One of the most important contributions of this study is the identification of how trend- and seasonality-adjusted relative volatility influences the forecasting performance of deep learning models. The interpretation of the results can be structured along the three research questions (RQ) outlined in the introduction and their associated theoretical frameworks.

### *5.1. Limits of market adaptation under extreme volatility (RQ1)*

The panel regression results revealed a significant positive relationship between trend- and seasonality-adjusted relative volatility (CV_STL) and the prediction error (MAPE) across all examined models. This mechanism represents a direct machine learning manifestation of the Adaptive Markets Hypothesis (AMH). During calm periods (2016–2018), deep learning models are able to effectively learn the stable patterns underlying price movements. However, during regime shifts triggered by exogenous shocks (COVID-19, geopolitical conflicts), the informational structure of the market suddenly becomes disrupted. The surge in relative volatility signals the invalidation of historical patterns, to which models are unable to adapt in real time, and therefore prediction error inevitably increases. The CV_STL indicator thus serves as an excellent proxy for the evolutionary pressure of the market and the failure of algorithmic adaptation.

### *5.2. The bias–variance trade-off and the trap of model complexity (RQ2)*

One of the most remarkable findings of the research is the difference in robustness among the examined architectures. Although the literature often favors more complex hybrid models, our results highlight the dangers of overfitting during turbulent periods. In crisis years, the prediction error of complex hybrid models (e.g., RNN-LSTM) increased drastically, often exceeding that of the base models. In contrast, GRU models operating with fewer gating mechanisms exhibited the lowest sensitivity to volatility (the smallest regression coefficient). This phenomenon represents a classic example of the bias–variance dilemma known from statistical learning theory. High-capacity hybrid models perform excellently during calm periods, but in crisis situations, when the signal level declines and stochastic noise (volatility) increases, the network tends to fit the noise, resulting in instability. In the case of GRU models, the principle of Occam's razor prevails: the tighter and simpler architecture enforces more regular learning, making the model more robust to noise.

### *5.3. Information set expansion and the curse of dimensionality (RQ3)*

The third critical finding is that multivariate (OHLC) models were, in most cases, more sensitive to relative volatility than their univariate counterparts. Although theoretically the inclusion of opening, high, and low prices provides additional information, during crisis periods this additional information often becomes a disadvantage. This can be explained by the curse of dimensionality and by the breakdown of correlation structures. In turbulent markets, the historical cross-correlations among OHLC data collapse (for example due to limit-price days or extreme intraday fluctuations). In such cases, the additional dimensions do not provide meaningful information but instead feed "multidimensional noise" into the network, amplifying error sensitivity. Thus, expanding the input structure in volatile environments may paradoxically increase the uncertainty of the system.

Overall, the research clearly demonstrates that relative volatility is one of the primary explanatory factors of prediction errors, and that the performance of deep learning models strongly depends on their ability to handle residual fluctuations that are free from trend and seasonality. The results indicate that during model selection attention should be paid not only to the average level of accuracy but also to its sensitivity to volatility. An important direction for future research may be the development of models that are specifically capable of adaptively handling residual volatility, whether through attention mechanisms or through the integration of regime-switching subsystems.

The interpretation of the results in the context of the Adaptive Markets Hypothesis and the bias–variance trade-off should be understood as indicative rather than confirmatory. The empirical analysis does not directly test these theoretical frameworks but provides patterns that are broadly consistent with their implications. Therefore, these interpretations are intended to offer a conceptual lens rather than definitive empirical validation.

### *5.4. Limitations*

The results of the research provide valuable insight into the relationship between trend- and seasonality-adjusted relative volatility and the performance of deep learning models, but several limitations must be considered when interpreting them. First, the applied STL decomposition is deterministic, so the separation of trend and seasonality can be sensitive to window sizes and smoothing parameters. Other decomposition techniques - for example, TBATS, Prophet, or machine learning–based component segmentation - may produce different residual structures, which can influence the magnitude of relative volatility. Second, although the study covers 12 different assets, the obtained results may not necessarily generalize to all financial markets, especially those with substantially different liquidity or microstructure. Furthermore, the definition of crises was done ex post, which necessarily simplifies the true complexity of market regimes. Third, while the model parameterization was uniform and controlled, it does not exhaust the full hyperparameter space. Certain configurations or regularization techniques (dropout, weight decay) could potentially improve the robustness of some models under high volatility. In addition, the input structure of the multivariate models could be further refined, as the present study tested only a basic variant. Finally, the study approaches relative volatility as a static measure and does not explicitly model its temporal dynamics. This may limit a deeper understanding of how changes in volatility affect the short-term adaptability of the models.

## 6. Conclusions and Future Directions

The study addressed a fundamental question: how machine learning models respond to structural shocks affecting financial markets and the resulting extreme volatility. Our findings clearly demonstrate that the accuracy of algorithmic forecasts depends not only on technical model settings but also on the macroeconomic and informational stability of the regime under investigation.

In line with the Adaptive Markets Hypothesis (AMH), it is demonstrated that an increase in trend- and seasonality-adjusted relative volatility (CV_STL) systematically reduces the reliability of forecasts, regardless of the asset examined. A primary conclusion of the research is that increasing model complexity does not represent an appropriate solution for handling turbulent periods. On the contrary, confirming the bias-variance trade-off, the findings demonstrate that hybrid and excessively deep architectures tend to internalize noise and suffer from overfitting during crisis periods. Among the examined architectures, GRU networks with relatively simpler gating mechanisms proved to be the most robust to volatility-induced shocks. Furthermore, it is highlighted that increasing the dimensionality of the input space (OHLC data) may amplify prediction errors during crisis periods, as the intraday correlation structures among price components tend to break down.

From a practical perspective, the study emphasizes that for financial institutions and risk managers, model selection should prioritize volatility sensitivity (robustness) rather than the "average" accuracy measured during calm market periods. In volatile market environments, simpler (Occam-type) univariate structures may provide greater reliability than parametrically overcomplicated systems. Consequently, future research and algorithmic development should focus on making objective functions more risk-sensitive and on stabilizing dedicated volatility-handling modules capable of filtering extreme noise (e.g., attention mechanisms).

Future research can follow several promising directions building on the results of the present study. One important avenue is the dynamic modelling of residual volatility, for example by integrating GARCH, FIGARCH, or Deep Volatility Networks. These approaches could reveal how the temporal evolution of volatility feeds back into the forecasting errors of deep learning models. It is also worth incorporating regime-identification methods, such as hidden Markov models, variational autoencoders, or clustering techniques, to automatically detect market states. Such models could adjust the network architecture or hyperparameters according to the detected regime, thereby increasing forecasting stability in turbulent periods. Regarding input data, another research direction could be the integration of macroeconomic or market-wide indicators, which may help contextualize changes in residual components. Similarly, experimenting with preprocessing "noisy" regressors or high-frequency filtering of the residual signal could be valuable. On the model side, transformer-based architectures and attention mechanisms represent a promising area, as they can potentially focus more effectively on the critical moments of residual volatility. Likewise, the development of hybrid models - for example, with volatility-based gating or adaptive network depth - could bring significant improvements. Finally, from a practical perspective, extending the current analysis to a real-time, rolling-window framework would better approximate the environment faced by market participants. This would allow an understanding of how quickly and to what extent the models can adapt to sudden volatility shocks.

**Author Contributions:** Conceptualization, L.V., T.T. and T.B.; methodology, L.V.; software, L.V.; validation, L.V., T.T. and T.B.; formal analysis, T.B.; investigation, T.T.; resources, L.V.; data curation, L.V.; writing-original draft preparation, L.V.; writing-review and editing, L.V.; visualization, L.V.; supervision, T.B.; project administration, T.T. All authors have read and agreed to the published version of the manuscript.

**Funding:** This research received no external funding.

**Statement on Data Availability:** The data used in this study are publicly available from Yahoo Finance (www.finance.yahoo.com) and CoinMarketCap (www.coinmarketcap.com).

**Conflicts of Interest:** The authors declare no conflicts of interest.

**Appendix A**

| Abbreviation | Full Term |
|---|---|
| AHM | Adaptive Markets Hypothesis |
| CLATT | Convolutional Neural Network - Bidirectional Long Short-term Memory – Attention Mechanism |
| CV_STL | Coefficient of Variation - Seasonal-Trend Decomposition using Loess |
| DL | Deep Learning |
| GRU | Gated Recurrent Unit |
| HMM | Hidden Markov Model |
| HMM-GRU | Hidden Markov Model - Gated Recurrent Unit |
| HMM-LSTM | Hidden Markov Model - Long Short-Term Memory |
| LSTM | Long Short-Term Memory |
| MAE | Mean Absolute Error |
| MAPE | Mean Absolute Percentage Error |
| ML | Machine Learning |
| ReLU | Rectified Linear Unit |
| RMSE | Root Mean Square Error |
| RNN | Recurrent Neural Network |
| STL | Seasonal-Trend Decomposition using Loess |
| Tanh | Hyperbolic Tangent (Activation Function) |
| TFT | Temporal Fusion Transformer |

## Appendix B

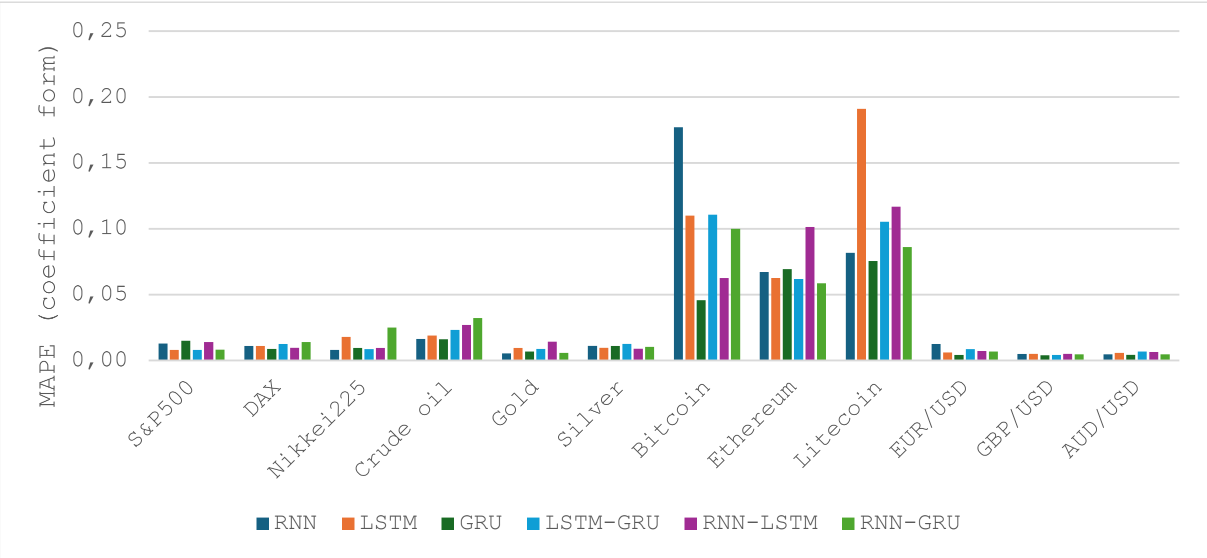


**Figure B1.** MAPE indicators estimated with univariate method between 1 January 2018 and 30 June 2018.
Source: own editing.

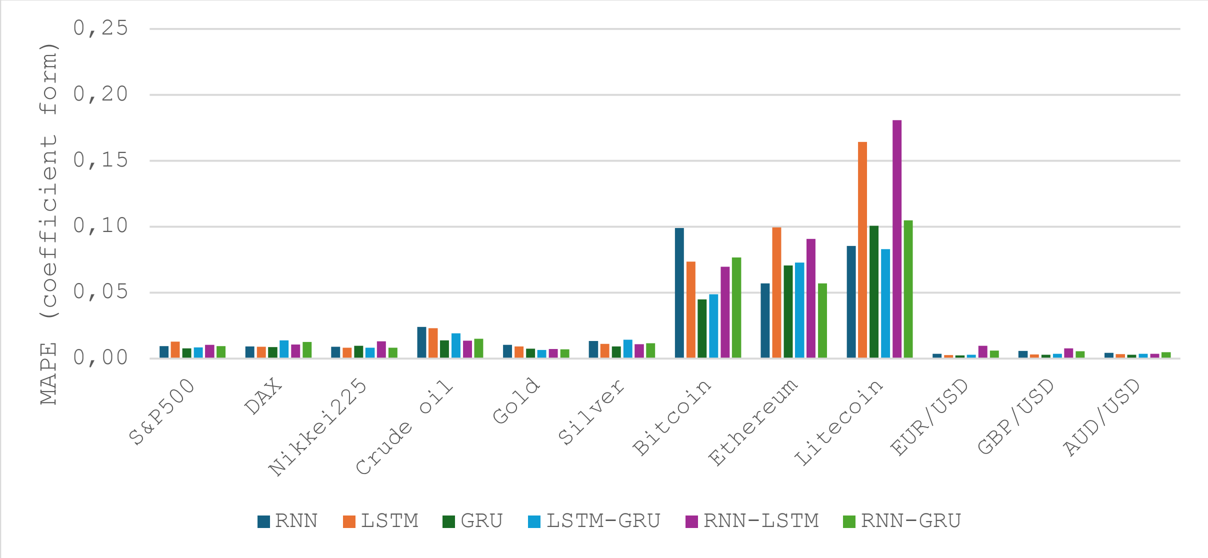


**Figure B2.** MAPE indicators estimated with multivariate method between 1 January 2018 and 30 June 2018.
Source: own editing

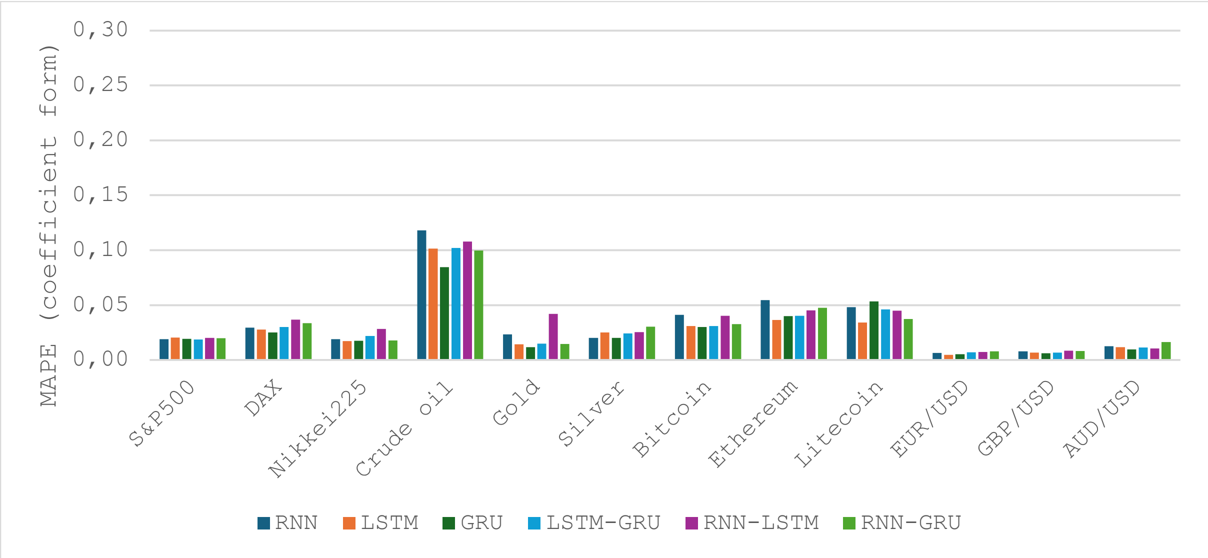


**Figure B3.** MAPE indicators estimated with univariate method between 1 January 2020 and 30 June 2020.
Source: own editing.

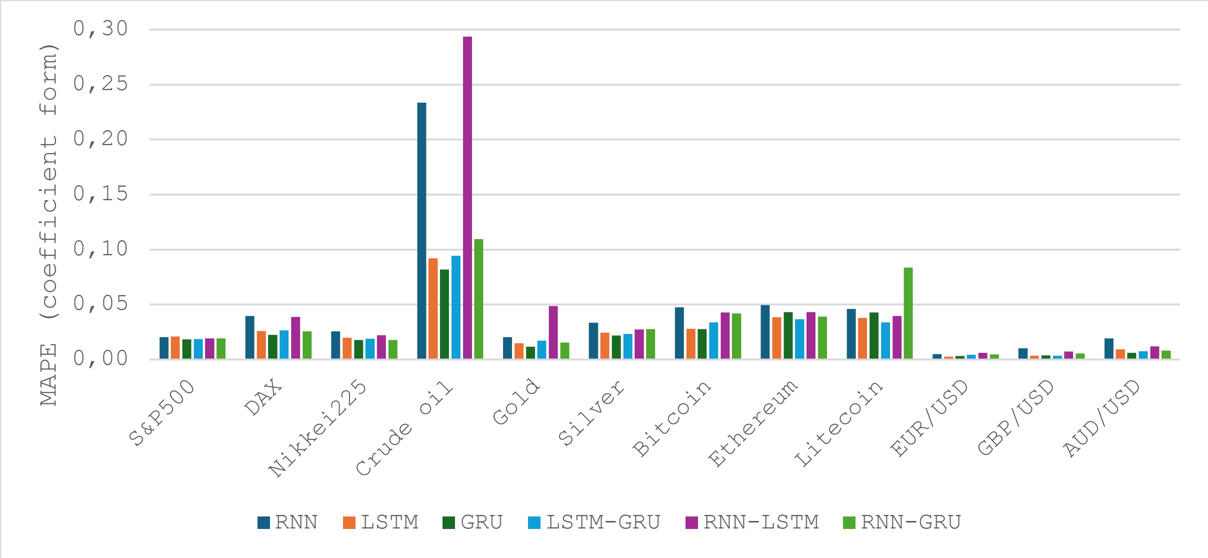


**Figure B4.** MAPE indicators estimated with multivariate method between 1 January 2020 and 30 June 2020.
Source: own editing.

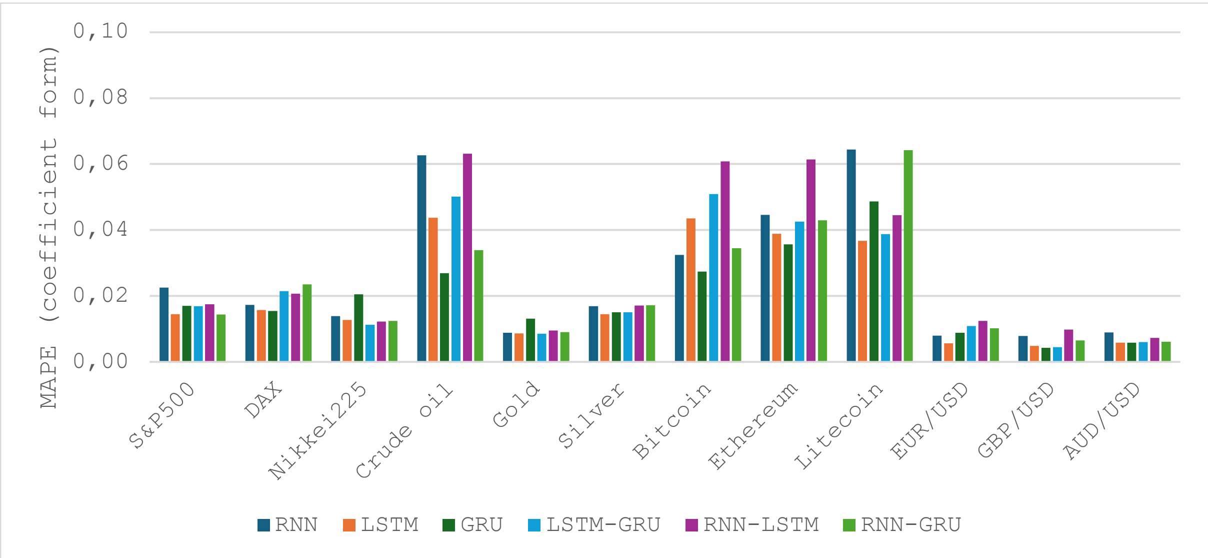


**Figure B5.** MAPE indicators estimated with univariate method between 1 January 2022 and 30 June 2022.
Source: own editing.

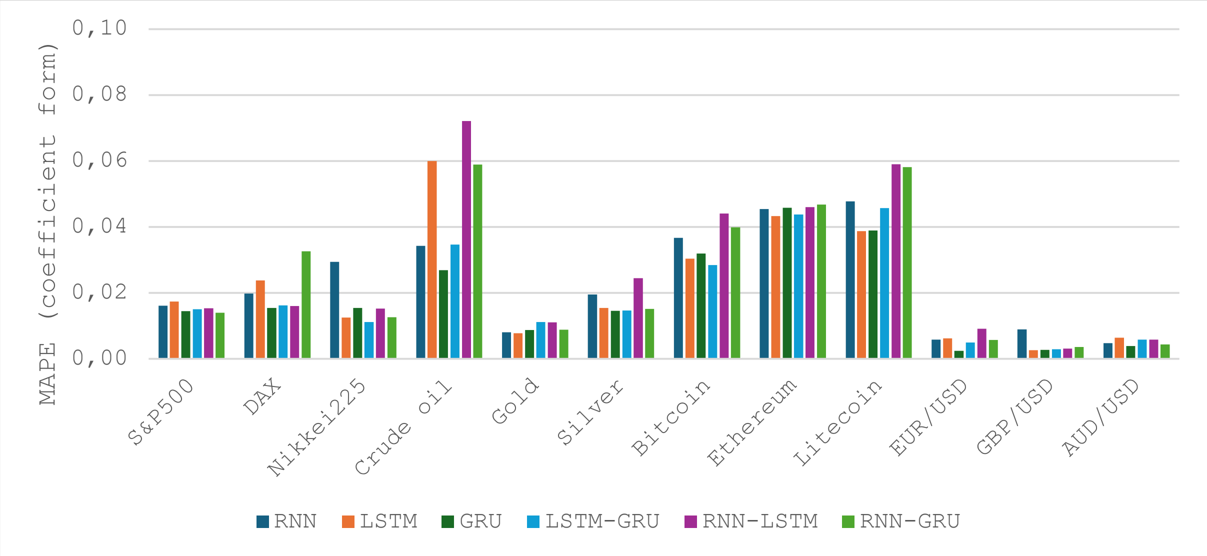


**Figure B6.** MAPE indicators estimated with multivariate method between 1 January 2022 and 30 June 2022.
Source: own editing.

## References

1. Mirza, F. K., Pekcan, Ö., Hekimoğlu, M., & Baykaş, T. (2025). Stock price forecasting through symbolic dynamics and state transition graphs with a convolutional recurrent neural network architecture. *Neural Computing and Applications*, *37*, 15855-15890. https://doi.org/10.1007/s00521-025-11325-z
2. Glova, A. M. G., & Barrios, E. B. (2025). Modelling Mixed-Frequency Time Series with Structural Change. *Computational Economics*, *65*, 3237-3258. https://doi.org/10.1007/s10614-024-10672-8
3. Pramod, A., Naicker, H. S., & Tyagi, A. K. (2021). Machine learning and deep learning: Open issues and future research directions for the next 10 years. In A. K. Tyagi (Ed.), *Computational analysis and deep learning for medical care: Principles, methods, and applications* (pp. 463-490). Wiley. https://doi.org/10.1002/9781119785750.ch18
4. Sakib, M., Mustajab, S., & Alam, M. (2025). Ensemble deep learning techniques for time series analysis: a comprehensive review, applications, open issues, challenges, and future directions. *Cluster Computing*, *28*, Article 73(2025). https://doi.org/10.1007/s10586-024-04684-0
5. Li, G., & Jung, J. J. (2023). Deep learning for anomaly detection in multivariate time series: Approaches, applications, and challenges. *Information Fusion*, *91*, 93-102. https://doi.org/10.1016/j.inffus.2022.10.008
6. Tissaoui, K., Boubaker, S., Hkiri, B., & Azibi, N. (2024). Multi-model transfer function approach tuned by PSO for predicting stock market implied volatility explained by uncertainty indexes. *Scientific Reports*, *14*, Article 22711. https://doi.org/10.1038/s41598-024-74456-8
7. Zhu, Y., Taasim, S. I., & Daud, A. (2025). Volatility modelling and tail risk estimation of financial assets: Evidence from gold, oil, Bitcoin, and stocks for selected markets. *Risks*, *13*(7), Article 138. https://doi.org/10.3390/risks13070138
8. Sivakumar, G. (2025). HMM-LSTM Fusion Model for Economic Forecasting. *arXiv preprint* arXiv:2501.02002. https://doi.org/10.48550/arXiv.2501.02002

9. El Zaar, A., Mansouri, A., Benaya, N., Bakir, T., & El Allati, A. (2025). Hybrid Transformer-CNN architecture for multivariate time series forecasting: Integrating attention mechanisms with convolutional feature extraction. *Journal of Intelligent Information Systems, 63*, 1233-1264. https://doi.org/10.1007/s10844-025-00937-5
10. Bhogade, V., & Nithya, B. (2024). Time series forecasting using transformer neural network. *International Journal of Computers and Applications, 46*(10), 880-888. https://doi.org/10.1080/1206212X.2024.2396321
11. Kafi, A., Rahman, M. M., Arefin, S M S., Sultana, N., Rafi, M.A., Zakaria, R.M. (2025). Forecasting currency volatility in global markets using transformer models: Implications for US trade and investment strategies. *Journal of Economics, Finance and Accounting Studies, 7*(5), 30-43. https://doi.org/10.32996/jefas.2025.7.5.4
12. Lo, A. W. (2004). The Adaptive Markets Hypothesis: Market efficiency from an evolutionary perspective. *The Journal of Portfolio Management, 30*, 15-29.
13. Hastie, T., Friedman, J. H., & Tibshirani, R. (2009). *The elements of statistical learning: Data mining, inference, and prediction.* Springer. https://doi.org/10.1007/978-0-387-21606-5
14. Zhang, X., & Zhang, T. (2022). Barrier option pricing under a Markov Regime switching diffusion model. *The Quarterly Review of Economics and Finance, 86*, 273-280. https://doi.org/10.1016/j.qref.2022.08.001
15. Carpinteyro, M., Venegas-Martínez, F., & Aali-Bujari, A. (2021). Modelling precious metal returns through fractional jump-diffusion processes combined with Markov regime-switching stochastic volatility. *Mathematics, 9*(4), Article 407. https://doi.org/10.3390/math9040407
16. Shi, Y. (2023). Long memory and regime switching in the stochastic volatility modelling. *Annals of Operations Research, 320*, 999-1020. https://doi.org/10.1007/s10479-020-03841-z
17. Avinash, G., Ramasubramanian, V., Paul, R. K., Ray, M., Dahiya, S., Iquebal, M. A., Godara, S., & Manjunatha, B. (2024). Price Forecasting of TOP (Tomato, Onion and Potato) Commodities using Hidden Markov-based Deep Learning Approach. *Statistics and Applications, 22*(2), 1-28.
18. Aydogan-Kilic, D., & Selcuk-Kestel, A. S. (2023). Modification of hybrid RNN-HMM model in asset pricing: univariate and multivariate cases. *Applied Intelligence, 53*, 23812-23833. https://doi.org/10.1007/s10489-023-04762-7
19. Mari, C., & Mari, E. (2023). Deep learning based regime-switching models of energy commodity prices. *Energy Systems, 14*, 913-934. https://doi.org/10.1007/s12667-022-00515-6
20. Mari, C., & Mari, E. (2025). Stochastic DNN-based models meet hidden Markov models: A challenge on natural gas prices at the Henry Hub. *Neural Computing and Applications*, *37*, 14997–15016. https://doi.org/10.1007/s00521-025-11286-3
21. Ouyang, Z. S., Yang, X. T., & Lai, Y. (2021). Systemic financial risk early warning of financial market in China using Attention-LSTM model. *The North American Journal of Economics and Finance, 56*, Article 101383. https://doi.org/10.1016/j.najef.2021.101383
22. Zhang, J., Ye, L., & Lai, Y. (2023). Stock price prediction using CNN-BiLSTM-Attention model. *Mathematics, 11*(9), Article 1985. https://doi.org/10.3390/math11091985
23. Luo, A., Zhong, L., Wang, J., Wang, Y., Li, S., & Tai, W. (2024). Short-term stock correlation forecasting based on CNN-BiLSTM enhanced by attention mechanism. *IEEE Access*, 12, 29617-29632. https://doi.org/10.1109/ACCESS.2024.3369419
24. Yu, H. (2025). An Adaptive Sharpe Ratio-Based Temporal Fusion Transformer for Financial Forecasting. *Frontiers in Humanities and Social Research, 2*(1), 54-65. https://doi.org/10.71465/fhsr23
25. Li, S., & Xu, S. (2025). Enhancing stock price prediction using GANs and transformer-based attention mechanisms. *Empirical Economics, 68*, 373-403. https://doi.org/10.1007/s00181-024-02644-6
26. Yahoo. (n.d.). *Yahoo Finance*. https://finance.yahoo.com/
27. CoinMarketCap. (n.d.). *CoinMarketCap*. https://coinmarketcap.com/
28. Bai, Y., Xie, J., Liu, C., Tao, Y., Zeng, B., & Li, C. (2021). Regression modelling for enterprise electricity consumption: A comparison of recurrent neural network and its variants. International *Journal of Electrical Power & Energy Systems, 126*, Artile 106612. https://doi.org/10.1016/j.ijepes.2020.106612
29. Xiao, H., Chen, Z., Cao, R., Cao, Y., Zhao, L., & Zhao, Y. (2022). Prediction of shield machine posture using the GRU algorithm with adaptive boosting: A case study of Chengdu Subway project. *Transportation Geotechnics, 37*, Article 100837. https://doi.org/10.1016/j.trgeo.2022.100837
30. Ortu, M., Uras, N., Conversano, C., Bartolucci, S., & Destefanis, G. (2022). On technical trading and social media indicators for cryptocurrency price classification through deep learning. *Expert Systems with Applications, 198*, Article 116804. https://doi.org/10.1016/j.eswa.2022.116804

31. Zhao, L., Li, Z., Qu, L., Zhang, J., & Teng, B. (2023). A hybrid VMD-LSTM/GRU model to predict non-stationary and irregular waves on the east coast of China. *Ocean Engineering*, *276*, Article 114136. https://doi.org/10.1016/j.oceaneng.2023.114136
32. Islam, M. S., & Hossain, E. (2021). Foreign exchange currency rate prediction using a GRU-LSTM hybrid network. *Soft Computing Letters, 3*, Article 100009. https://doi.org/10.1016/j.socl.2020.100009
33. Faru, S. H., Waititu, A., & Nderu, L. (2023). A Hybrid Neural Network Model Based on Transfer Learning for Forecasting Forex Market. *Journal of Data Analysis and Information Processing*, *11*(2), 103-120. https://doi.org/10.4236/jdaip.2023.112007
34. Cleveland, R. B., Cleveland, W. S., McRae, J. E., & Terpenning, I. (1990). STL: A seasonal-trend decomposition procedure based on loess. *Journal of Official Statistics*, *6*(1), 3–73.
35. He, H., Gao, S., Jin, T., Sato, S., & Zhang, X. (2021). A seasonal-trend decomposition-based dendritic neuron model for financial time series prediction. *Applied Soft Computing*, *108*, Article 107488. https://doi.org/10.1016/j.asoc.2021.107488
36. Brockwell, P. J., & Davis, R. A. (2016). *Introduction to time series and forecasting.* Springer.