STATISTICS, OPTIMIZATION AND INFORMATION COMPUTING
Stat., Optim. Inf. Comput., Vol. **x**, Month 202x, pp 0–18.
Published online in International Academic Press (www.IAPress.org)

Advancing Volatility Forecasting in Financial Indices: Integrating GARCH Models, Multifractal Indicators, and Deep Learning

Naima Oualla*, Mohammed Salah Chiadmi, Youssef Lamrani Alaoui

Mohammed V University in Rabat, IFELAB-LERMA, Mohammadia School of Engineering, Rabat, MOROCCO

Abstract Accurate volatility prediction is essential for effective investment strategies and risk awareness. Yet, the intricate and ever-changing characteristics of markets pose considerable challenges, motivating the use of hybrid frameworks by integrating heteroscedastic models, multifractal analysis, and deep learning techniques. While heteroscedastic models are simple and widely adopted, they often fail to reflect the inherent nonlinearities and multifractal properties of volatility. In contrast, LSTM, GRU, and Transformers, while capable of capturing complex structures, require well-chosen explanatory variables to deliver accurate forecasts.
Accordingly, this study conducts a rigorous comparative investigation across the Dow Jones Islamic Market Index, the Dow Jones Global Index, and the S&P 500. We confirm the existence of multifractal scaling and evaluate the performance of deep learning models based on historical features against hybrid models integrating GARCH-type forecasts and multifractal indicators. Results demonstrate that integrating GARCH, EGARCH, and FIGARCH features significantly improves accuracy by embedding key stylized facts such as volatility clustering, asymmetry, and long memory, with statistical significance confirmed by the Diebold-Mariano test. Furthermore, findings indicate that while standalone multifractal features are insufficient, they serve as complementary inputs. Rather than proposing a single novel model, the contribution of this work lies in a systematic analysis of feature complementarity, demonstrating that guiding deep learning with econometric signals enhances predictive robustness across diverse market structures.

Keywords Volatility forecasting, heteroscedastic models, deep learning, multifractality, hybrid models

AMS 2010 subject classifications 62M10, 91B84

DOI: 10.19139/soic-2310-5070-3018

1. Introduction

The increasing complexity of global financial markets has made forecasting volatility dynamics a fundamental challenge in both academic research and practical financial applications, given the chaotic, uncertain, and incomplete nature of market information [43] . This challenge is particularly pronounced for the Dow Jones Islamic Market Index (DJIMI), which has gained significant attention due to its growing role in global finance and the distinctive features of participative investments [24]. To provide a comprehensive perspective on these dynamics and benchmark the results against conventional markets, this study extends the analysis to include the Dow Jones Global Index (DJG) and the Standard & Poor's 500 (S&P 500). This comparative approach allows for a robust examination of volatility structures across different investment paradigms.

The Generalized Autoregressive Conditional Heteroscedastic (GARCH) model, introduced by Bollerslev [4], and its exponential extension EGARCH by Nelson [26], have proven effective in capturing time-varying volatility and asymmetries. However, these models typically assume a short memory process, where the influence of shocks decays exponentially over time. In contrast, empirical studies have shown that financial volatility often exhibits

*Correspondence to: Naima Oualla (Email: naimaoualla2001@gmail.com). Laboratories for study and research in applied mathematics (LERMA), Mohammadia School of Engineering, Mohammed V University, Rabat, Morocco.

ISSN 2310-5070 (online) ISSN 2311-004X (print)
Copyright © 202x International Academic Press

long memory behavior, with persistent autocorrelation (see [11]). The Fractionally Integrated GARCH (FIGARCH) model, proposed by Baillie et al. [2], introduces a fractional differencing parameter that adjusts the model's memory depth. This extension allows for a more accurate representation of persistent volatility patterns. Nevertheless, due to the complicated non-linear correlation structures between variables, along with the massive volume of data to be analyzed, the volatility prediction results of these GARCH-type models remain insufficient [17].

Deep learning models such as Recurrent Neural Networks (RNNs), particularly Long Short-Term Memory (LSTM) and Gated Recurrent Unit (GRU) have shown strong performance in volatility prediction, as they take into consideration nonlinearities and temporal dependencies [13]. Despite the success of RNN variants, their inherent sequential processing nature imposes limitations on capturing long-range dependencies and prevents parallelization during training. As the sequence length increases, the ability of LSTMs and GRUs to retain information from earlier time steps can diminish, creating a bottleneck for modeling global temporal structures.

To address these limitations, we introduce the Transformer architecture into our comparative framework. Transformers rely entirely on self-attention mechanisms, allowing the model to weigh the significance of different time steps simultaneously—regardless of their distance in the past—thereby capturing complex, non-linear volatility dynamics more effectively than sequential models. However, they rely heavily on rich and informative explanatory variables that reflect the empirical structure of financial markets. Exploring various types of input data in deep learning models helps identify which individual inputs or combinations thereof are most effective [9]. The focus of this paper is to extract the most meaningful inputs; to do so, we turned to the multifractal analysis.

Originally developed in physics, the concept of multifractality has been successfully applied to finance; studies have confirmed the performance of multifractal analysis in reflecting the complex aspect of markets [35]. Among the most prominent tools, Multifractal Detrended Fluctuation Analysis (MF-DFA), developed by Kantelhardt et al. [18], has been integrated as a dataset enrichment tool, making it possible to extract fine-grained characteristics, thereby providing a better understanding of market behavior. Its integration into deep neural networks is expected to improve predictive performance and provide richer explanations for models, particularly in contexts where market dynamics are complex and multivariate [10].

Prior research has detected the presence of multifractal scaling structures stock indices [25, 31], highlighting the need for more robust analyses of volatility dynamics. In our study, we also identified multifractal behavior, which we incorporated as an explanatory variable to optimize the predictive performance.

First, we use LSTM, GRU, and Transformers, trained solely on past realized volatility and past returns, to establish a baseline performance. Second, we enhance them by adding volatility forecasts from GARCH-type models (GARCH, EGARCH, and FIGARCH). Third, we perform a Multifractal Detrended Fluctuation Analysis (MF-DFA) to extract the generalized Hurst exponent h, a key indicator in characterizing the temporal dynamics of volatility. Values of $h < 0.5$ suggest anti-persistent behavior, where increases are likely followed by decreases and vice versa, indicating a tendency toward mean-reversion. Conversely, $h > 0.5$ implies persistent behavior, where trends are more likely to continue. When $h = 0.5$, the time series behaves as a purely random process [18]. Thus, this exponent provides a compact measure of the underlying memory and structure, making it valuable in volatility forecasting. Finally, we develop a comprehensive hybrid model integrating all these explanatory features—realized volatility, returns, GARCH-type forecasts, and multifractal indicators to assess their combined effect on predictive performance.

The article is organized as follows: Section 2 presents a review of existing studies related to time series dynamics and volatility prediction. Then, Section 3 introduces the methodological framework, focusing on heteroscedastic models, multifractal analysis approaches, and predictive models. Section 4 describes the dataset, including its statistical properties and distributional characteristics, which guide the choice of appropriate predictive models.

Section 5 explores the multifractal structure of the time series to understand the evolution of its scaling behavior and extract relevant indicators that are later used as input in deep learning models. Section 6 is devoted to experiments and results, where we detail the predictive models, input variables, model architectures, and evaluation metrics, as well as the obtained results.

The concluding section summarizes the key results, discusses the benefits and shortcomings of every model, and proposes directions for future research.

2. Literature Review

Volatility forecasting has been predominantly grounded in autoregressive conditional heteroskedasticity (ARCH) models, introduced by Engle [12]. Bollerslev [4] extended this framework by proposing the Generalized ARCH (GARCH) model, which became the foundation for much of the volatility modeling literature. These models estimate conditional variance based on past squared residuals and variances, and have proven effective in capturing volatility clustering in financial time series [4, 15]. Building on this foundation, numerous extensions were developed in response to specific shortcomings of the standard GARCH framework [5].

One major limitation of the model is its failure to incorporate the leverage effect—also referred to as the asymmetric volatility response—whereby adverse shocks exert a more pronounced influence on market volatility than favorable shocks. As empirical studies have confirmed the presence of such asymmetries in financial time series, the EGARCH model was proposed by Nelson [26] to capture this behavior.

Moreover, several studies have highlighted the presence of persistent temporal structure, a feature that standard GARCH models fail to model. This gave rise to the FIGARCH model proposed by Baillie et al. [2], which incorporates a fractional differencing parameter to model the memory depth.

In parallel, volatility prediction has advanced considerably through machine learning algorithms. This area has emerged as a prominent research trend over the past decade, with deep learning approaches markedly enhancing predictive performance [9]. Neural network-based techniques, in particular, have demonstrated strong predictive capabilities in various studies (see [27,42]). In fact, LSTM architectures are widely regarded as essential components in numerous financial time series forecasting models. Among their variants, GRU networks have gained significant attention and demonstrated comparable performance to LSTM models.

However, despite their design intended to manage long-term dependencies, recurrent architectures are not immune to computational limitations. When processing extraordinarily long temporal sequences, LSTMs and GRUs can still suffer from gradient instability—specifically the vanishing or exploding gradient problem—which hampers their ability to retain information over extended horizons. By contrast, the Transformer architecture learns information from all positions in the entire sequence simultaneously, rather than sequentially. This parallel processing capability effectively avoids the information loss often encountered during iterative training. Consequently, a growing number of academics are turning to attention-based models to capture complex temporal dynamics that traditional recurrent networks fail to preserve [36].

Even with these advances, several challenges, such as limited interpretability and substantial data requirements, continue to affect the effectiveness of the model. In addition, the problem of delayed prediction has been widely documented; it refers to the delay between the occurrence of relevant market events and the model's ability to forecast their effects on time. Such delays result in forecasts that often lag behind actual market trends, limiting their ability to provide truly forward-looking insights. This issue is particularly critical given that financial markets respond almost instantaneously to new information and dynamic changes [9].

In light of these challenges, researchers have explored the incorporation of additional features that better characterize market behavior, leading to the development of hybrid models that seek to combine the strengths of deep learning and econometric models. For instance, Amirshahi and Lahmiri, in their research [1], developed hybrid models in which GARCH-type volatility forecasts were used as inputs to a Deep Feedforward Neural Network (DFNN) and Long Short-Term Memory (LSTM) network, enhancing predictive performance on cryptocurrency markets. They concluded that GARCH-type volatility is an informative variable that improves the precision of predictions. In the same vein, Michańków et al. [22] proposed a hybrid model combining GARCH and GRU and evaluated its robustness across three distinct datasets: the S&P 500 index, gold prices, and Bitcoin. Their results demonstrated that such hybrid architectures are more performant than standalone models. Following the paradigm shift towards self-attention mechanisms, recent studies have investigated integrating GARCH features into Transformer-based architectures to overcome the sequential limitations of RNNs. Pioneering this approach, Ramos-Pérez et al. [30] introduced a "Multi-Transformer" framework that merges varying GARCH-type forecasts with an attention-based network. Their study on the S&P 500 volatility demonstrated that this architecture—specifically when enhanced by a bagging mechanism within the attention heads—statistically outperforms both standalone GARCH models and hybrid LSTM-GARCH configurations. Extending this line of inquiry, Mishra et al. [23]

validated the robustness of these Multi-Transformer hybrids across a broader spectrum of assets, including foreign exchange rates and individual stocks. Their empirical evidence highlights that Transformer-based hybrids maintain superior predictive accuracy not only during stable periods but also during regimes of extreme market turbulence, such as the COVID-19 pandemic, thereby confirming that econometric volatility inputs provide crucial guidance to the self-attention mechanism in noisy financial environments. Nevertheless, financial time series are highly sensitive and can be influenced by minor indicators, making the problem of selecting the most informative explanatory variables a persistent challenge [9].

More advanced statistical models have turned to fractal and multifractal modeling. The application of fractals in finance was notably introduced by Mandelbrot [21], whose work paved the way for multifractal modeling in asset returns. Building on Mandelbrot's pioneering work, Calvet and Fisher extend the multifractal paradigm by providing both theoretical insights and practical tools for analyzing and modeling volatility [8]. Kantelhardt, in his work [18], proposed a Multifractal Detrended Fluctuation Analysis (MF-DFA) as a tool to extract meaningful features. In particular, the generalized Hurst exponents which have proven effective in capturing both persistence and anti-persistence within time series. Multifractal behavior arises when the degree of temporal dependence is scale-dependent, varying across time horizons, in contrast to monofractal processes that exhibit uniform scaling [10].

A considerable number of studies used the MF-DFA to detect multifractality and its sources [7, 31, 39]. They suggest that the formation of multifractality is largely driven by nonlinear dependencies over time, the emergence of heavy-tailed events, and the departure of return distributions from normality. However, these studies mainly consider the MF-DFA as a descriptive tool for analyzing the dynamics and structure of financial markets, rather than as a method for extracting characteristics to train deep learning models for forecasting purposes. Nevertheless, only a limited number of studies have employed MF-DFA within a predictive framework. For instance, Ying Yuan et al. [41] found that incorporating multifractal characteristics into the modeling process significantly improves forecasting performance, especially in times of financial stress or turbulence. Poongjin Cho et al. [10], in their research, trained RNNs with Hurst exponents. They concluded that those multifractal indicators enhance the predictive ability of RNNs. Xiang Yu et al. [40] proposed a model that relies on multifractal analysis to detect complicated volatility features, and they combined this approach with an attention mechanism. More recently, Florindo et al. [14] advanced this integration by proposing GHENet, a model that explicitly injects non-linear dynamic information via generalized Hurst exponents into a deep learning predictor. A key innovation of their architecture is the use of a self-attention module specifically designed to process these fractal features, allowing the model to dynamically attend to the most relevant scaling behaviors. Their empirical results on worldwide stock indices demonstrate that this attention-based processing of Hurst exponents significantly boosts forecasting performance compared to traditional feature integration methods. While these models achieved superior prediction compared with standalone models, they provided only a limited exploration of the predictive power of different feature sets. In contrast, the present study extends the literature by integrating econometric, deep learning, and multifractal perspectives into a unified hybrid predictive framework. This work not only highlights the relative importance of diverse features but also demonstrates that combining them allows for capturing time-varying volatility, nonlinear dynamics, and multi-scale dependencies simultaneously.

3. Materials and Methods

3.1. GARCH Models

Although financial return series are generally uncorrelated over time, they exhibit volatility clustering, i.e., large (small) residuals tend to be followed by other large (small) residuals. This implies that while the conditional mean can often be treated as constant, the conditional variance displays temporal dependence [4]. To capture this feature, GARCH model proposed by Bollerslev [4] expresses the conditional variance as a linear function of its own lags.

Formally, the return process is written as the mean equation

$$r_t = \mu_t + \varepsilon_t, \tag{1}$$

where μ_t denotes the conditional mean and $\varepsilon_t = \sigma_t z_t$ is the heteroscedastic error term. The equation of the GARCH model is given by

$$\sigma_t^2 = \omega + \alpha(L)\varepsilon_t^2 + \beta(L)\sigma_t^2 \tag{2}$$

Given that L is a lag or backshift operator, this equation can also be written in expanded form as:

$$\sigma_t^2 = \omega + \sum_{i=1}^{p} \alpha_i \varepsilon_{t-i}^2 + \sum_{j=1}^{q} \beta_j \sigma_{t-j}^2. \tag{3}$$

Where p (respectively q) refers to the number of lagged σ_t^2 (respectively ε_t^2), with $\omega > 0$, $\alpha_i \geq 0$, and $\beta_j \geq 0$.

To account for the asymmetric response of volatility to shocks—often neglected by standard GARCH—the EGARCH framework proposed by Nelson [26] is applied, through the inclusion of a parameter γ.

$$\log \sigma_t^2 = \omega + \alpha(L)\left(\frac{\varepsilon_t}{\sigma_t}\right) + \gamma(L)\left(\left|\frac{\varepsilon_t}{\sigma_t}\right| - \mathbb{E}\left|\frac{\varepsilon_t}{\sigma_t}\right|\right) + \beta(L)\log \sigma_t^2 \tag{4}$$

Equivalently,

$$\log \sigma_t^2 = \omega + \sum_{i=1}^{p}\left[\alpha_i \frac{\varepsilon_{t-i}}{\sigma_{t-i}} + \gamma_i\left(\left|\frac{\varepsilon_{t-i}}{\sigma_{t-i}}\right| - \mathbb{E}\left|\frac{\varepsilon_{t-i}}{\sigma_{t-i}}\right|\right)\right] + \sum_{j=1}^{q} \beta_j \log \sigma_{t-j}^2. \tag{5}$$

We employ the Fractionally Integrated GARCH (FIGARCH) model, which extends the GARCH framework by introducing a fractional differencing parameter $d \in (0, 1)$ [2]. This parameter allows for a more flexible memory structure that better captures the long-term persistence observed in real-world financial markets.

Following the same notation used in our GARCH specification, the FIGARCH(p, d, q) model can be expressed as:

$$\sigma_t^2 = \omega[1 - \beta(1)]^{-1} + \lambda(L)\epsilon_t^2 \tag{6}$$

Where $\lambda(L) = \sum_{k=1}^{\infty} \lambda_k L^k$ is a lag polynomial with positive coefficients that define the infinite memory in the conditional variance, which is a distinctive feature of the FIGARCH model [2]. The model can also be reformulated as:

$$\sigma_t^2 = \frac{\omega}{1 - \sum_{j=1}^{q} \beta_j} + \sum_{k=1}^{\infty} \lambda_k \varepsilon_{t-k}^2. \tag{7}$$

3.2. Multifractal Detrended Fluctuation Analysis

The Detrended Fluctuation Analysis (DFA) method [28] has gained wide popularity for characterizing fractal scaling properties and uncovering extended correlation structures. Kantelhardt et al. [18] have extended this method by introducing the Multifractal Detrended Fluctuation Analysis (MF-DFA), which has become one of the most widely used techniques for extracting meaningful indicators that provide a means of evaluating the persistence within temporal sequences [10].

Let $z(t)_{t\in\mathbf{N}}$ be a time series of length n. The returns are defined as:

$$r_t = \left|\ln\left(\frac{z(t+\Delta t)}{z(t)}\right)\right|, \tag{8}$$

Where Δ_t denotes the discrete time step.

The analysis procedure is described as follows:

- **Step 1:** Define the accumulated profile y :

$$y(i) = \sum_{t=1}^{i}(r_t - \bar{r}), \quad i = 1, \ldots, n, \tag{9}$$

where $\bar{r}$ denotes the mean of the entire return series.

- **Step 2:** The profile is split into $n_s = \lfloor n/s \rfloor$ equally sized segments of length s, ensuring that they do not overlap. When n is not perfectly divisible by s, a residual portion of the profile remains unsegmented. To maintain statistical reliability, the segmentation procedure is reapplied from the opposite end, yielding a total of $2n_s$ segments.
- **Step 3:** Estimate the total trend using a polynomial regression. The variance of the residuals from these regressions is then computed as follows:

$$F^2(l,s) = \begin{cases} \dfrac{1}{s}\displaystyle\sum_{k=1}^{s}\left\{y\left[(l-1)s+k\right] - f_l^{(m)}(k)\right\}^2, & \text{Where } l = 1,\ldots,n_s \\ \dfrac{1}{s}\displaystyle\sum_{k=1}^{s}\left\{y\left[n-(l-n_s)s+k\right] - f_l^{(m)}(k)\right\}^2, & \text{Where } l = n_s+1,\ldots,2n_s \end{cases} \tag{10}$$

  where $f_l^{(m)}(k)$ is a polynomial of order m, used to approximate the local trends. In this study, we set the polynomial order to $m = 2$, as previous research indicates that in typical cases, this value provides an optimal choice [34].
- **Step 4:** By computing the mean, the fluctuation function is obtained as follows :

$$F_q(s) = \left\{\frac{1}{2n_s}\sum_{l=1}^{2n_s}\left[F^2(l,s)\right]^{q/2}\right\}^{1/q}. \tag{11}$$

  $q = 2$ corresponds to the standard DFA (see [28]).
- **Step 5:** Verify that the fluctuation function scales as a power law:

$$F_q(s) \propto s^{h(q)}, \tag{12}$$

  for sufficiently large s.
  This is done by analyzing the plots of $F_q(s)$ versus s, i.e., $\log(F_q(s)) = h(q)\log(s) + \text{constant}$.
  When $q \to 0$, the value of $F_0(s)$ is obtained by logarithmic averaging:

$$F_0(s) = \exp\left\{\frac{1}{4n_s}\sum_{l=1}^{2n_s}\ln\left[F^2(l,s)\right]\right\} \propto s^{h(0)}. \tag{13}$$

To establish the relationship between the Hurst exponent and the scaling exponent—another key indicator of multifractality—we proceed as follows:

We suppose that the series r_t is stationary, strictly positive, and normalized, that is, $r_t \geq 0$ and $\sum_{t=1}^{n} r_t = 1$. In this case, the detrending step can be omitted, as there are no underlying trends that require removal. This assumption leads to the following relation:

$$\left\{\frac{1}{2n_s}\sum_{l=1}^{2n_s}|y(ls) - y((l-1)s)|^q\right\}^{1/q} \propto s^{h(q)}, \tag{14}$$

If n is an exact multiple of s, i.e., $n_s = \frac{n}{s}$, then the series can be divided perfectly into n_s, then we obtain :

$$\sum_{l=1}^{2n_s}|y(ls) - y((l-1)s)|^q \propto s^{qh(q)-1}. \tag{15}$$

The term $y(ls) - y((l-1)s)$ is the sum of r_t within segment l of size s:

$$p_s(l) = \sum_{t=(l-1)s+1}^{ls} r_t = y(ls) - y((l-1)s). \tag{16}$$

The scaling exponent $\tau(q)$ is defined through the partition function $Z_q(s)$:

$$Z_q(s) = \sum_{l=1}^{n/s} |p_s(l)|^q \propto s^{\tau(q)}, \tag{17}$$

From equations (14), (15), and (16) we derive the following relation:

$$\tau(q) = qh(q) - 1. \tag{18}$$

It serves as a diagnostic for identifying multifractality: A linear relationship between $\tau(q)$ and q reflects a monofractal structure, whereas a nonlinear relationship reveals the presence of multifractality in the system.

3.3. The Structure of Predictive Models

The objective of this research is to predict the one-day-ahead realized volatility; each predictive model incorporates a specific combination of explanatory features to test their individual and joint contributions to volatility forecasting. The input variables vary across models and include past volatility, past returns, volatility forecasts from GARCH models, and multifractal indicators. The models compared are (i) deep learning models with historical features, (ii) hybrid GARCH–deep learning models, (iii) deep learning models based on multifractal indicators, and (iv) a final hybrid model that combines all the aforementioned features.

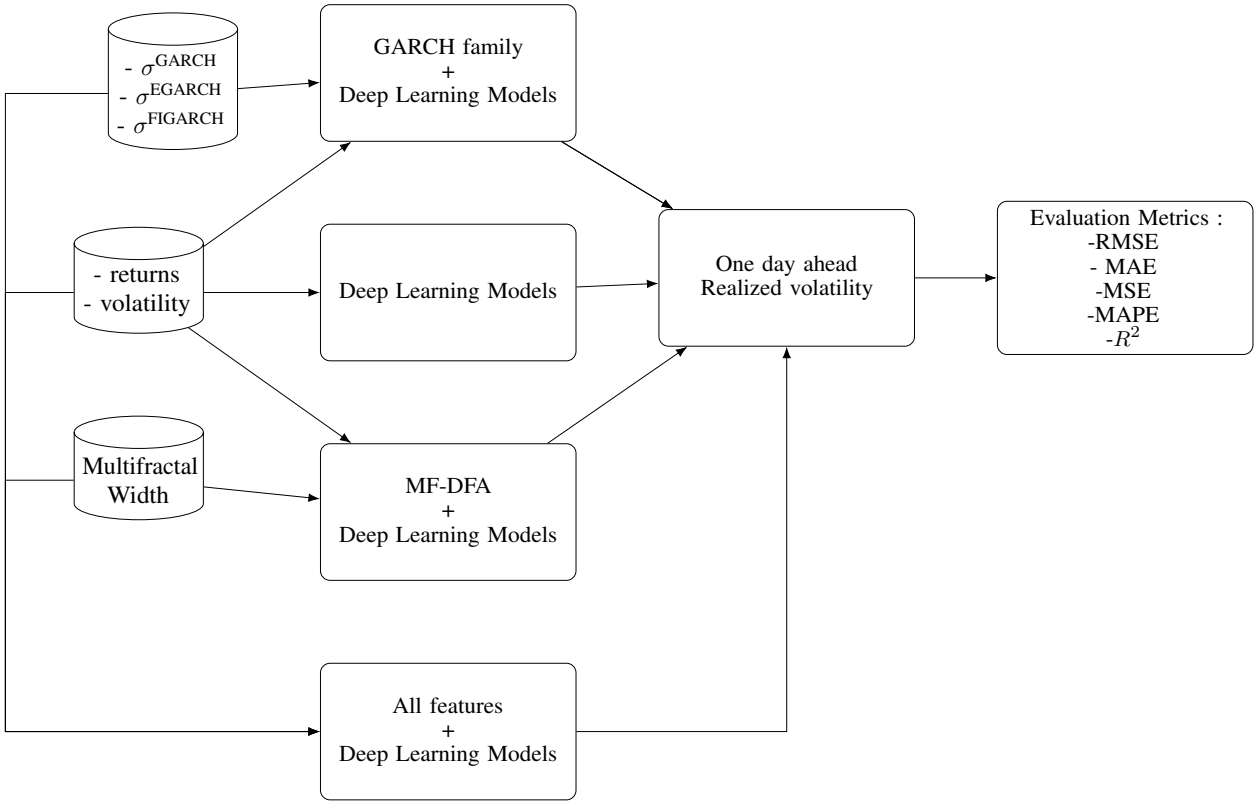


Figure 1. Architecture of the proposed predictive models.

4. Data Description

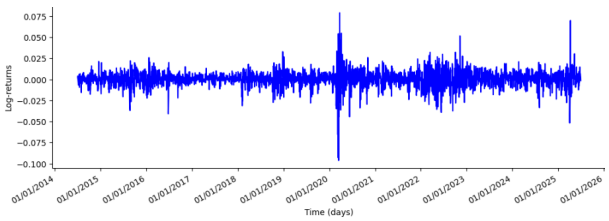
To ensure the robustness and generalizability of the proposed framework, the empirical analysis is conducted on three distinct stock market indices, analyzed over the period from July 1, 2014, to July 1, 2025. The primary focus is

the Dow Jones Islamic Market Index, which tracks stocks selected based on rules-driven criteria ensuring compliance with ethical investment principles[†], and includes 10 economic sectors, as presented in Table 1. As a representative benchmark for the global conventional market, we include the Dow Jones Global Index. Additionally, the Standard & Poor's 500 is incorporated to capture the dynamics of the U.S. large-cap equity market.

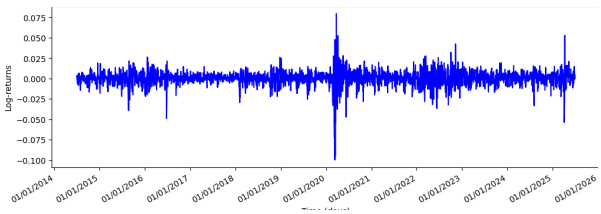
Table 1. Sectors of DJIMI

Sector Name
Basic Materials
Consumer Goods
Consumer Services
Financials
Health Care
Industrials
Oil & Gas
Technology
Telecommunications
Utilities

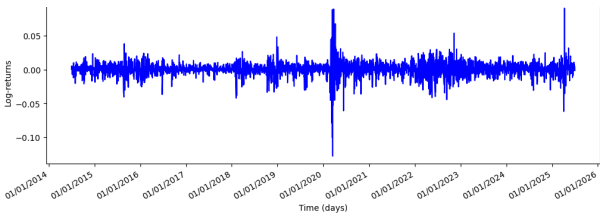
We define the daily log-return as $(r_t = \ln(z_t) - \ln(z_{t-1}))$, where z_t represents the daily closing value of the market index. The temporal evolution of returns for the three indices is illustrated in (Fig. 2). These plots visually confirm that volatility is time-varying and far from constant. Notably, the phenomenon of volatility clustering is clearly observable across all series: periods of relative tranquility are intermittently interrupted by bursts of high turbulence, where large fluctuations tend to be followed by other large fluctuations.This behavior aligns with the well-known empirical findings on financial stylized facts reported by Cont [11].



(a) Dow Jones Islamic Index



(b) Dow Jones Global Index



(c) S&P 500

Figure 2. Daily returns of the Dow Jones Islamic Market Index, Dow Jones Global Index, and S&P 500

[†]https://www.spglobal.com/

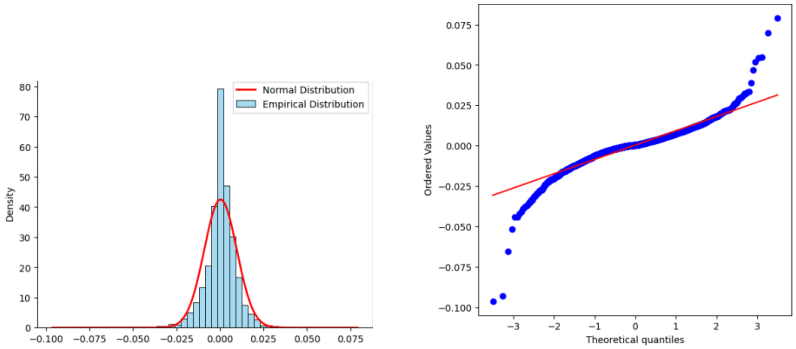


Figure 3. DJIMI returns distribution

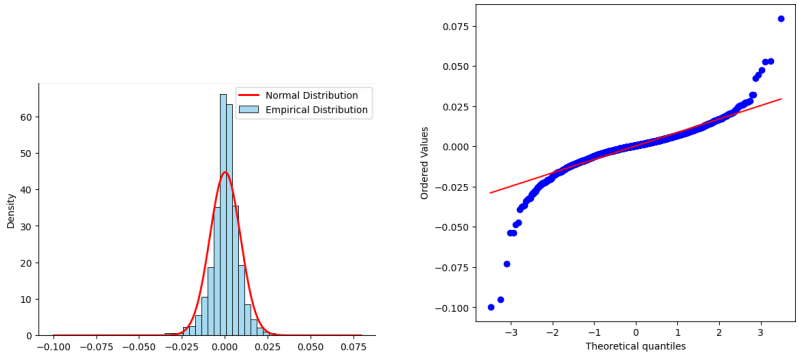


Figure 4. Global Dow Jones Index returns distribution

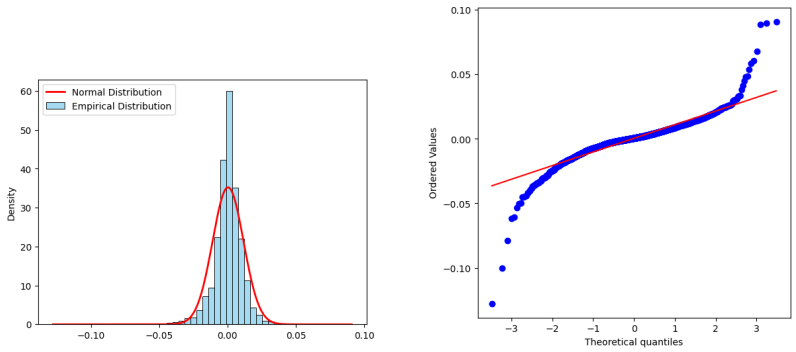


Figure 5. SP&500 Index returns distribution

Table 2. Descriptive statistics of returns

Index	**Mean**	**standard deviation**	**Skewness**	**Kurtosis**	**ARCH-LM (p-val)**	**Lo's test(p-val)**
DJIMI	0.000310	0.009391	-0.683	13.069	8.48×10^{-169}	0.2605
DJG	0.000254	0.008922	-1.122	17.052	3.05×10^{-178}	0.2742
S&P 500	0.000414	0.011328	-0.643	15.768	4.12×10^{-184}	0.2813

The study demonstrates that the returns for all three indices deviate from normality, as evidenced by their skewness and kurtosis (see Figures 3, 4, and 5). Moreover, strong evidence of heteroscedasticity was detected across all series using the ARCH-LM test[‡], with a p_{value} significantly lower than the 0.05 threshold. This result validates the hypothesis of heteroscedasticity and confirms the fact that the variance changes as a function of time, suggesting the suitability of ARCH-type models. Additionally, Lo's modified R/S test[§] was employed to assess the persistence in the time series. The test fails to detect statistically significant persistence in any of the indices ($p_{value} \geq 0.05$) (see Table 2). These preliminary findings support our choice of models and help to justify the strengths and limitations of the different modeling approaches considered in this study.

5. Multifractal Properties of the Studied Financial Indices

Following the MF-DFA methodology, the functions $F_q(s)$ are evaluated over a range of temporal scales s, starting from 16 days and extending up to $\frac{n}{4}$, where n represents the total number of observations. The choice of the upper bound $\frac{n}{4}$ is motivated by the findings in [18], which highlight that, for very large scales (i.e., $s \geq \frac{n}{4}$), the number of available non-overlapping segments, given by $n_s = \frac{n}{s}$, becomes too limited. This small number of segments undermines the statistical robustness of the averaging process involved in computing $F_q(s)$. Consequently, scales beyond this limit are excluded from the regression step used to estimate the generalized scaling exponent $h(q)$.

The analysis uses a range of q values from -10 to 10, with steps of 2. As reported in [18], positive q values highlight large fluctuations, while negative q values give more weight to small variations. This difference helps to detect how the behavior of the time series changes between calm and volatile periods, giving a better understanding of market dynamics.

If the time series under study exhibits monofractal characteristics, the function $h(q)$ remains constant across all q values, indicating a uniform scaling across all fluctuation sizes. However, when $h(q)$ varies with q, this signals that fluctuations of different magnitudes scale differently — a core feature of multifractal systems.

This multifractal nature is further supported by the plots of $F_q(s)$ versus s illustrated in Figure 6. A distinct linear scaling behavior is observed consistently across all three indices (DJIMI, DJ Global, and S&P 500). This linearity confirms that the power-law relation holds true regardless of the specific market structure, validating the applicability of the MF-DFA method for this comparative study.

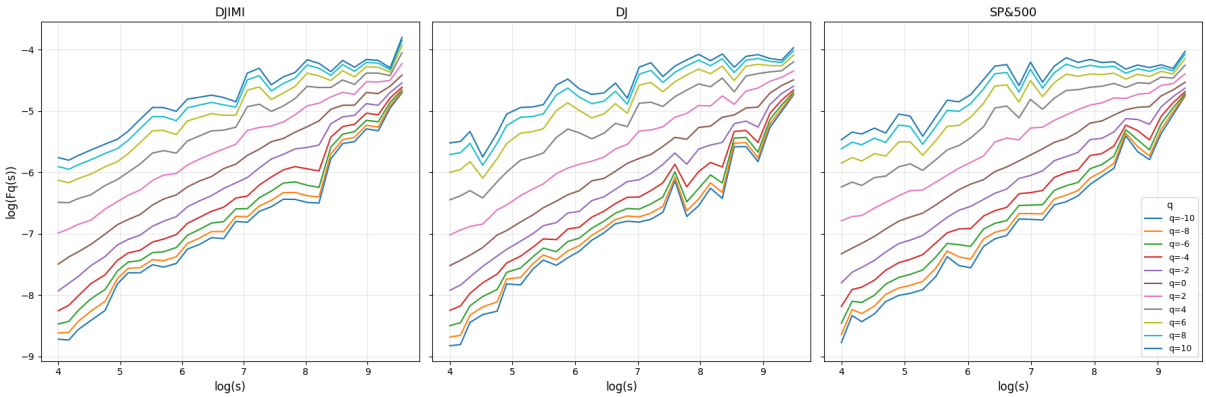


Figure 6. $F(q)$ as function of scales for different values of q

[‡]This test checks whether the variance changes over time, helping to detect ARCH-type behavior.
[§]Lo's modified R/S test is an extension of the classical rescaled range analysis, designed to detect long memory in time series

As q increases from -10 to 10, the estimated $h(q)$ for all three indices exhibits a distinct decreasing trend, dropping from high values to lower values. Such variation in $h(q)$ values reflects heterogeneous market behaviors across scales, underscoring the relevance of multifractal-based features for volatility modeling.

Additionally, the mass exponent function defined in equation (18) also exhibits a strong non-linear relationship with q, which serves as another robust indicator of multifractality. In contrast, for monofractal series, where $h(q)$ is constant, and $\tau(q)$ varies linearly with q (see Figure 7).

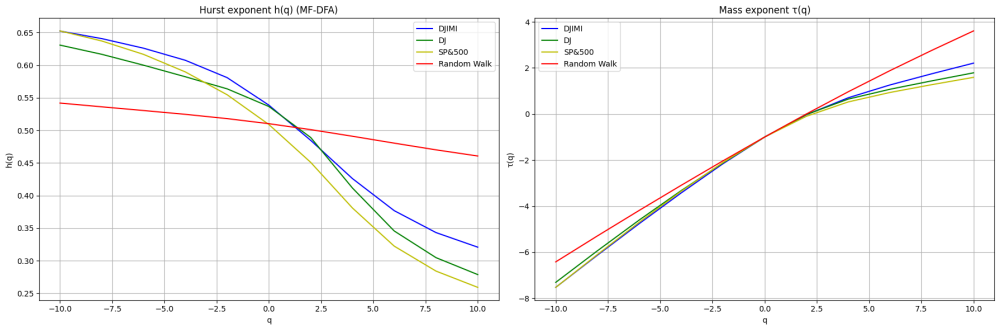


Figure 7. Generalized Hurst Exponent and Mass Function as functions of q.

6. Experiments and Results

The realized volatility is defined as the empirical rolling volatility calculated as the standard deviation of daily returns over a fixed-size window of $w = 30$ days. Formally, it is expressed as:

$$\sigma_t = \sqrt{\frac{1}{w}\sum_{i=0}^{w-1}(r_{t-i} - \bar{r}_t)^2}, \quad \text{where} \quad r_t = \ln\left(\frac{z_t}{z_{t-1}}\right)$$

where r_t denotes the log-return at time t and $\bar{r}_t$ is the mean return over the interval $[t - w + 1, t]$. Realized volatility is directly observable from historical data, which makes it a particularly relevant and practical target for forecasting and for supporting informed investment strategies [33].

We consider a wide range of models, from traditional econometric approaches to deep learning and hybrid architectures. Each type of model uses specific input features, as described below, in the following sections:

6.1. Proposed Architectures and Feature Sets

Deep Learning Models for Volatility Forecasting: To capture the complex temporal dependencies of realized volatility, we employ three distinct Deep Learning architectures: LSTM, GRU, and Transformers. While LSTM and GRU networks rely on recurrent mechanisms to process sequential data, the Transformer architecture leverages a self-attention mechanism—specifically implemented here with 4 attention heads—to capture long-range dependencies in parallel. Despite their architectural differences, we adopt a unified data-driven approach where those baseline models ingest the exact same input vector. This vector is constructed using an optimal lag length L determined via grid search, specifically ranging from 5 to 30 past observations. This choice is justified by the structure of financial markets, it captures dynamics from day traders operating on a weekly basis (short-term) to investors focusing on monthly trends (medium-term) [6]. The complete input vector at time t, shared across all baseline models, is thus defined as:

$$X_t^{baseline} = [\sigma_{t-L}, \ldots, \sigma_{t-1},\ r_{t-L}, \ldots, r_{t-1}]$$

where $L \in \{5, 30\}$.

GARCH–deep learning Hybrid Architectures for Volatility Prediction: To accurately predict realized volatility in a way that reflects market reality, it is essential to develop a model that accounts for various empirical characteristics. In this regard, we incorporate GARCH-type forecasts into the deep learning input. Consistent with the prevailing literature, we utilize the $(1, 1)$ order, which has been repeatedly proven to be sufficient for modeling financial volatility while avoiding the overfitting risk associated with higher-order parameters [1, 17].

Formally, the extended input vector at time t is given by:

$$X_t^{hybrid} = \left[\sigma_{t-L}, \ldots, \sigma_{t-1}, r_{t-L}, \ldots, r_{t-1}, \sigma_t^{GARCH-type}\right]$$

where $\sigma_t^{GARCH-type}$ represents the volatility forecast generated by the GARCH, EGARCH, or FIGARCH model [1, 17].

This allows the model to capture statistical structure from GARCH-type models and nonlinear patterns from recent data. Hybrid models that combine deep learning and GARCH may improve the robustness and interpretability of volatility forecasts, but do not account for the multifractal nature of financial time series.

Integration of MF-DFA-Based Indicators into Deep Learning Architectures: To enrich the model with multifractal information while mitigating the risk of overfitting due to high dimensionality, we adopt a selective feature engineering approach. The generalized Hurst exponents $h(q)$ are computed using the MF-DFA method applied to a moving window of $w = 126$ trading days. This window size is chosen to strike a balance between statistical robustness and the ability to capture evolving market dynamics. This choice is supported by the recent work of Florindo et al. [14], who established that a window of 100 trading days represents the minimum threshold to ensure computational stability, noting that larger values increase computational cost without improving forecasting performance.

Furthermore, instead of incorporating the entire spectrum of q values as inputs, we focus on the extreme moments $q = -5$ and $q = 5$. Several studies have highlighted that these specific values effectively characterize the scaling behavior of small and large fluctuations, respectively [16, 19]. Consequently, we construct a compact multifractal feature, denoted as Δh_t, which represents the multifractal spectrum width (or degree of multifractality) defined as:

$$\Delta h_t = h_t(q_{min}) - h_t(q_{max}) \tag{19}$$

where $q_{min} = -5$ and $q_{max} = 5$. This indicator captures the richness of the multifractal structure. By using this condensed metric, we preserve the essential multifractal information without increasing the computational complexity of the deep learning models. The input vector is then defined as:

$$X_t = [\sigma_{t-L}, \ldots, \sigma_{t-1}, r_{t-L}, \ldots, r_{t-1}, \Delta h_{t-L}, \ldots, \Delta h_{t-1}] \tag{20}$$

Deep Learning Models with Combined Explanatory Features: In this section, we propose a fully integrated forecasting framework that combines multiple sources of information to enhance the predictive power of Deep Learning Models. Specifically, we use a set of explanatory variables including historical data, volatility forecasts obtained from GARCH-type models (GARCH, EGARCH, FIGARCH), and the multifractal feature.

This hybrid design reflects the hypothesis of complementarity, which is based on the idea that each type of input feature captures distinct but relevant aspects of the volatility structure. Historical returns and realized volatility reflect the recent price behavior and short-term market dynamics. GARCH-type forecasts, on the other hand, embed statistical information such as conditional heteroscedasticity, leverage effects, and long-memory behavior that are not always detectable through raw data. Finally, Multifractal spectrum derived from MF-DFA offers insights into the multiscale structure of the series. The input vector is then given by:

$$X_t = \left[\sigma_{t-L}, \ldots, \sigma_{t-1},\ r_{t-L}, \ldots, r_{t-1}, \sigma^{GARCH-type},\ \Delta h_{t-L}, \ldots, \Delta h_{t-1}\right]$$

6.2. Experimental Setup and Evaluation

To ensure a rigorous evaluation and prevent look-ahead bias, we adopted a strict temporal data splitting strategy. The total dataset was first divided into a training set (first 80%) and a hold-out test set (last 20%). The test set was strictly preserved for the final out-of-sample evaluation and was never used during the training or hyperparameter tuning phases.

During the grid search optimization, the training set was further split: 80% was used for training and 20% was set aside as a validation set. This validation subset played a crucial role in monitoring model convergence and tuning hyperparameters.

Regarding the training dynamics, we set a maximum of 250 epochs. However, to avoid overfitting and reduce computational cost, we implemented an early stopping mechanism. This callback monitored the validation loss and stopped training if no improvement was observed after a patience period (10 epochs), automatically restoring the model weights that achieved the lowest validation error.

The hyperparameters were optimized via grid search to identify the configuration that minimizes the Mean Squared Error. We simultaneously optimized structural parameters (number of layers, neurons) and input configurations (lag lengths for returns, volatility, and multifractal width). The specific search space explored is detailed in Table 3. All models were trained using the Adam optimizer with a fixed learning rate of 0.001 and a batch size of 32, ensuring stable convergence.

To ensure the robustness and reproducibility of our findings, and to minimize the bias induced by random seeding, each optimal model configuration was executed for 30 independent runs. Consequently, all evaluation metrics reported in the subsequent results section represent the average performance across these 30 trials.

Table 3. Hyperparameter search space and training settings.

Hyperparameter	**Search Space / Value**
Architecture & Training	
Hidden Layers	$[1, 3]$
Neurons per Layer	$[64, 128]$
Batch Size	32
Learning Rate	0.001
Max Epochs	250 (with Early Stopping)
Optimizer	Adam
Input Structure (Lags)	
Returns Lag	$[5, 30]$
Volatility Lag	$[5, 30]$
Multifractal Lag	$[5, 30]$

The forecasting accuracy of the proposed models is evaluated using five standard metrics that complement each other. Here, σ_i denotes the observed volatility, $\hat{\sigma}_i$ the predicted volatility, $\bar{\sigma}_i$ the mean of observed values, and n the number of observations.

Table 4. Evaluation metrics and their descriptions

Evaluation Metrics	Description
$\text{MAE} = \frac{1}{n}\sum_{i=1}^{n}\lvert\sigma_i - \hat{\sigma}_i\rvert$	The mean absolute error measures the average absolute deviation between predicted and actual volatility, providing an overall measure of accuracy.
$\text{MSE} = \frac{1}{n}\sum_{i=1}^{n}(\sigma_i - \hat{\sigma}_i)^2$	The mean squared error emphasizes larger deviations by squaring the errors, making it sensitive to significant prediction errors.
$\text{RMSE} = \sqrt{\frac{1}{n}\sum_{i=1}^{n}(\sigma_i - \hat{\sigma}_i)^2}$	The root mean squared error offers an interpretable metric in the same scale as the original data, summarizing the overall forecasting error.
$\text{MAPE} = \frac{1}{n}\sum_{i=1}^{n}\frac{\lvert\sigma_i - \hat{\sigma}_i\rvert}{\lvert\sigma_i\rvert}$	The mean absolute percentage error evaluates prediction accuracy in relative terms by expressing the absolute error as a percentage of the actual values, making it particularly useful for comparing forecast performance.
$R^2 = 1 - \frac{\sum_{i=1}^{n}(\sigma_i - \hat{\sigma}_i)^2}{\sum_{i=1}^{n}(\sigma_i - \bar{\sigma}_i)^2}$	The coefficient of determination evaluates the explanatory power of the model, indicating how well the predicted values approximate the observed data. An R^2 value close to 1 indicates that the model explains most of the variability in the observed data.

6.3. Results and Discussion

Tables below summarize the forecasting performance of all models using five evaluation metrics.

Table 5. Comparison of forecasting performance (DJIMI Index)

Model	MSE	RMSE	MAE	R^2	MAPE
LSTM	1.07×10^{-6}	0.001031	0.000592	0.9137	6.88
LSTM_MF	1.44×10^{-6}	0.001191	0.000745	0.8847	8.86
GARCH_LSTM	9.36×10^{-7}	0.000963	0.000587	0.9243	6.67
EGARCH_LSTM	1.11×10^{-6}	0.001044	0.000645	0.9099	7.20
FIGARCH_LSTM	1.18×10^{-6}	0.001071	0.000680	0.9049	7.68
ALL_LSTM	1.22×10^{-6}	0.001098	0.000657	0.9022	7.26
GRU	9.39×10^{-7}	0.000964	0.000597	0.9240	6.90
GRU_MF	1.15×10^{-6}	0.001069	0.000657	0.9079	7.78
GARCH_GRU	1.23×10^{-6}	0.001092	0.000738	0.9008	8.34
EGARCH_GRU	8.58×10^{-7}	0.000913	0.000584	0.9306	6.90
FIGARCH_GRU	9.81×10^{-7}	0.000982	0.000651	0.9206	7.61
ALL_GRU	8.73×10^{-7}	0.000930	0.000575	0.9302	6.84
Transformer	9.38×10^{-7}	0.000952	0.000609	0.9250	7.29
Transformer_MF	3.29×10^{-6}	0.001504	0.000935	0.7370	10.79
GARCH_Transformer	1.18×10^{-6}	0.001030	0.000656	0.9060	7.85
EGARCH_Transformer	1.52×10^{-6}	0.001026	0.000640	0.8781	7.31
FIGARCH_Transformer	9.01×10^{-7}	0.000927	0.000572	0.9279	6.77
ALL_Transformer	1.02×10^{-6}	0.000980	0.000629	0.9184	7.61

Table 6. Comparison of forecasting performance (Dow Jones Index)

Model	MSE	RMSE	MAE	R^2	MAPE
LSTM	9.92×10^{-7}	0.000994	0.000549	0.8971	7.36
LSTM_MF	1.17×10^{-6}	0.001077	0.000640	0.8806	9.10
GARCH_LSTM	7.54×10^{-7}	0.000867	0.000530	0.9218	7.14
EGARCH_LSTM	7.77×10^{-7}	0.000873	0.000545	0.9194	7.46
FIGARCH_LSTM	9.28×10^{-7}	0.000959	0.000550	0.9037	7.16
ALL_LSTM	8.81×10^{-7}	0.000926	0.000604	0.9099	8.80
GRU	9.46×10^{-7}	0.000967	0.000574	0.9018	7.66
GRU_MF	1.09×10^{-6}	0.001044	0.000606	0.8882	8.52
GARCH_GRU	9.12×10^{-7}	0.000950	0.000556	0.9054	7.24
EGARCH_GRU	1.05×10^{-6}	0.001001	0.000638	0.8907	8.27
FIGARCH_GRU	7.79×10^{-7}	0.000877	0.000549	0.9192	7.46
ALL_GRU	9.77×10^{-7}	0.000975	0.000636	0.9001	9.27
Transformer	9.28×10^{-7}	0.000937	0.000573	0.9052	7.85
Transformer_MF	1.04×10^{-6}	0.001016	0.000652	0.8932	9.00
GARCH_Transformer	7.49×10^{-7}	0.000859	0.000499	0.9234	6.77
EGARCH_Transformer	1.29×10^{-6}	0.000958	0.000620	0.8679	8.42
FIGARCH_Transformer	2.40×10^{-6}	0.001231	0.000717	0.7544	9.34
ALL_Transformer	8.50×10^{-7}	0.000872	0.000544	0.9130	7.53

Table 7. Comparison of forecasting performance (SP500 Index)

Model	MSE	RMSE	MAE	R^2	MAPE
LSTM	1.75×10^{-6}	0.001321	0.000742	0.9195	7.63
LSTM_MF	2.04×10^{-6}	0.001421	0.000895	0.9074	9.75
GARCH_LSTM	1.30×10^{-6}	0.001137	0.000680	0.9403	7.15
EGARCH_LSTM	1.71×10^{-6}	0.001301	0.000713	0.9214	7.18
FIGARCH_LSTM	1.53×10^{-6}	0.001233	0.000693	0.9296	6.96
ALL_LSTM	2.18×10^{-6}	0.001467	0.000824	0.9007	8.05
GRU	1.75×10^{-6}	0.001316	0.000742	0.9195	7.74
GRU_MF	1.95×10^{-6}	0.001392	0.000804	0.9114	8.75
GARCH_GRU	1.65×10^{-6}	0.001271	0.000771	0.9243	7.74
EGARCH_GRU	1.37×10^{-6}	0.001160	0.000709	0.9368	7.66
FIGARCH_GRU	1.23×10^{-6}	0.001104	0.000666	0.9436	7.29
ALL_GRU	1.55×10^{-6}	0.001233	0.000781	0.9295	8.45
Transformer	4.11×10^{-6}	0.001834	0.001163	0.8128	12.53
Transformer_MF	5.16×10^{-6}	0.002034	0.001283	0.7653	13.60
GARCH_Transformer	1.62×10^{-6}	0.001240	0.000700	0.9263	7.20
EGARCH_Transformer	2.54×10^{-6}	0.001394	0.000944	0.8844	10.43
FIGARCH_Transformer	2.19×10^{-6}	0.001350	0.000758	0.9002	7.58
ALL_Transformer	3.17×10^{-6}	0.001636	0.001053	0.8558	11.02

The forecasting performance of the considered models across the DJIMI, DJ Global, and S&P 500 indices is summarized in Tables 5, 6, and 7, respectively. The comparison relies on five distinct metrics.

The baseline models (LSTM, GRU, and Transformer), trained solely on historical returns and realized volatility, establish a robust benchmark but exhibit varying degrees of effectiveness across markets. As shown in Table 5 (DJIMI), the LSTM and GRU models achieve an RMSE of approximately 0.0010, indicating reasonable predictive

capability. However, the Transformer architecture generally exhibits higher error rates in standalone configurations, particularly for the Dow Jones Global index (Table 6), where it records the highest RMSE (0.001834). This performance discrepancy aligns with the comparative analysis of Bilokon and Qiu [3], who investigated the applicability of Transformers in financial forecasting. Their study concluded that while Transformers may show advantages in predicting absolute price sequences, LSTM-based models demonstrate superior and more robust performance on difference sequences, such as price movements and returns. Consequently, without the guidance of domain-specific features, the self-attention mechanism appears less effective than recurrent architectures in capturing the short-term, stationary dynamics of realized volatility.

The integration of econometric volatility forecasts constitutes the most significant performance driver. Across all three indices, models augmented with GARCH, EGARCH, or FIGARCH inputs consistently reduce prediction errors. For instance, in the case of the DJIMI (Table 5), the EGARCH-GRU model achieves the lowest RMSE (0.000913), significantly outperforming the standalone GRU (0.000964). Similarly, for the S&P 500 (Table 7), the GARCH-Transformer yields the best overall performance with an RMSE of 0.000859, improving upon the baseline Transformer (0.000937). This result is also supported by other research, as the hybridization of GARCH-type models and deep learning has generally shown better performance compared to standalone models [1].

A more nuanced picture emerges regarding the contribution of multifractal features. Models relying solely on multifractal indicators frequently exhibit higher error metrics than the baselines. This finding contrasts with conclusions drawn in recent literature [38, 40], where multifractal indicators were reported to significantly enhance standalone predictive performance. This divergence can be justified by the specific stochastic properties of the datasets analyzed. As reported in Table 2, Lo's Modified R/S test failed to detect statistically significant long-memory persistence in the DJIMI, DJ Global, and S&P 500 indices. Since the predictive power of generalized Hurst exponents Δh_t relies heavily on the presence of persistent autocorrelation structures, their standalone efficacy is naturally diminished in markets exhibiting weak memory. However, their value becomes evident in the fully integrated "ALL" architectures. In the DJ Global index, the ALL-GRU model achieves a competitive RMSE of 0.001233 with a high R^2 of 0.9295. This demonstrates that while multifractal exponents lack the directional signal required for precise one-step-ahead forecasting when used in isolation, they act as effective complementary variables.

Table 8. Diebold-Mariano Test Statistics (Positive values indicate Hybrid > Baseline)

Model Variation	**DJIMI**	**DJ Global**	**S&P 500**
Panel A: Comparison against Baseline LSTM			
LSTM-MF	-1.19	-4.40^{***}	-0.84
GARCH-LSTM	1.77^{*}	2.88^{***}	3.51^{***}
EGARCH-LSTM	0.14	3.53^{***}	0.99
FIGARCH-LSTM	-0.22	1.63	1.36
ALL-LSTM	-0.45	2.35^{**}	-0.96
Panel B: Comparison against Baseline GRU			
GRU-MF	-2.94^{***}	-3.39^{***}	-3.82^{***}
GARCH-GRU	-2.58^{***}	0.68	0.22
EGARCH-GRU	3.38^{***}	-0.29	2.96^{***}
FIGARCH-GRU	0.14	3.06^{***}	3.06^{***}
ALL-GRU	2.07^{**}	1.13	0.92
Panel C: Comparison against Baseline Transformer			
Trans-MF	-7.33^{***}	-4.59^{***}	-0.41
GARCH-Trans	2.17^{**}	1.36	4.25^{***}
EGARCH-Trans	3.66^{***}	2.75^{***}	4.99^{***}
FIGARCH-Trans	3.03^{***}	0.61	4.28^{***}
ALL-Trans	1.86^{*}	1.78^{*}	0.86

Note: Values represent the DM test statistic.
***, **, and * denote significance at 1%, 5%, and 10% levels.

While the reduction in error metrics points towards the superiority of hybrid architectures, we employ the Diebold-Mariano (DM) test to verify the statistical significance of these improvements. Table 8 presents the pairwise comparison of selected hybrid models against their respective baselines.

The results of the Diebold-Mariano test, presented in Table 8, provide a rigorous statistical validation of the proposed hybrid frameworks across the three indices. The test statistics reveal distinct patterns regarding the contribution of each feature set across different architectures.

Consistent with the RMSE findings, the DM test confirms that integrating multifractal features in isolation generally fails to yield statistically significant improvements over the baselines. This inefficiency is particularly pronounced for the DJIMI index, where the *Trans-MF* configuration records the lowest test statistic in the study ($DM = -7.33^{***}$), indicating a severe deterioration compared to the standalone Transformer. Similarly, negative significant values are observed for the *LSTM-MF* on DJ Global ($DM = -4.40^{***}$). This reinforces the conclusion that while multifractal exponents capture structural roughness, they lack the directional volatility signal required for precise short-term forecasting when used without concurrent variance estimators.

In contrast, the integration of econometric volatility forecasts delivers the most consistent statistical gains, although the optimal configuration depends on both the backbone architecture and the specific market. Within the LSTM framework, improvements remain moderate for the DJIMI (with *GARCH-LSTM* showing marginal significance at 1.77^{*}), whereas the *EGARCH-LSTM* model proves robust on the DJ Global (3.53^{***}). For the GRU network, results are highly index-dependent: while the *GARCH-GRU* underperforms on the Islamic index, the *EGARCH-GRU* configuration emerges as a top performer for the DJIMI (3.38^{***}), suggesting that the asymmetric leverage effect is a critical driver for this specific market. Conversely, the *FIGARCH-GRU* excels on the DJ Global (3.06^{***}), indicating a better capture of long-memory features in that context. Finally, the Transformer baseline benefits most dramatically from hybridization across all boards; notably, the *EGARCH-Trans* model achieves massive improvements for both the DJIMI (3.66^{***}) and S&P 500 (4.99^{***}), as the explicit volatility path bridges the gap in measuring serial dependence for the non-recurrent architecture.

Regarding the comprehensive models combining all features, the results generally confirm they outperform the baselines (e.g., *ALL-GRU* on DJIMI: 2.07^{**} and *ALL-LSTM* on DJ Global: 2.35^{**}). However, they do not systematically generate the highest t-statistics compared to the simpler GARCH or EGARCH hybrids.

7. Conclusion and Future Work

This study proposed and rigorously evaluated a novel hybrid forecasting framework that integrates econometric volatility estimates (GARCH family) and multifractal features (MF-DFA) into advanced Deep Learning architectures (LSTM, GRU, and Transformers). By applying these models to the Dow Jones Islamic Market Index, the Dow Jones Global Index, and the S&P 500, we assessed their ability to capture the complex, non-linear, and long-memory dynamics of financial volatility.

The empirical results lead to three major conclusions. First, the integration of econometric volatility forecasts significantly enhances the predictive accuracy of deep learning models, validating the synergy between domain-specific econometric theory and data-driven representation learning. The Transformer architecture benefited the most from this hybridization, with the EGARCH-Transformer achieving the highest statistical significance in forecasting improvements across the S&P 500 and the DJIMI. This suggests that providing an explicit volatility path effectively compensates for the Transformer's lack of inherent recurrence, allowing its self-attention mechanism to focus on refining the volatility trajectory.

Second, standalone multifractal features are insufficient for short-term forecasting. Models relying solely on multifractal inputs frequently failed to outperform the baselines. This finding implies that while generalized Hurst exponents successfully characterize structural market roughness, they lack the directional signal required for one-step-ahead prediction unless paired with robust variance estimators.

Third, model performance is highly sensitive to the specific characteristics of the target index. For the Islamic market, models incorporating asymmetric volatility yielded the most robust gains (e.g., EGARCH-GRU), highlighting the importance of capturing the leverage effect in Shariah-compliant equities. Conversely, the Dow

Jones Global Index favored models capable of handling long-memory dependencies, such as the FIGARCH-GRU. Furthermore, while the comprehensive "ALL" models generally outperformed the baselines, they consistently yielded lower test statistics than the GARCH-Deep Learning hybrids.

In summary, this research validates the hypothesis of feature complementarity, demonstrating that the integration of domain-specific econometric signals significantly enhances the predictive robustness of deep learning architectures. Future research could extend this framework by exploring high-frequency data to better capture intraday multifractality or by integrating the recently proposed attention-based processing of Hurst exponents to better exploit fractal dynamics. Additionally, testing these hybrid architectures on other volatile asset classes, such as cryptocurrencies or commodities, remains a promising avenue for validation.

REFERENCES

1. B. Amirshahi and S. Lahmiri, *Hybrid deep learning and GARCH-family models for forecasting volatility of cryptocurrencies*, Machine Learning with Applications, vol. 12, pp. 100465, 2023.
2. R. T. Baillie, T. Bollerslev, and H. O. Mikkelsen, *Fractionally integrated generalized autoregressive conditional heteroscedasticity*, Journal of Econometrics, vol. 74, no. 1, pp. 3–30, 1996.
3. P. Bilokon and Y. Qiu, *Transformers versus LSTMs for electronic trading*, arXiv preprint arXiv:2309.11400, 2023.
4. T. Bollerslev, *A conditionally heteroscedastic time series model for speculative prices and rates of return*, The Review of Economics and Statistics, pp. 542–547, 1987.
5. C. Brooks, *Introductory econometrics for finance*, Cambridge University Press, 2014.
6. B. Korkusuz, *Beyond the S&P 500: examining the role of external volatilities in market forecasting*, Review of Economic Design, vol. 29, no. 4, pp. 767–794, 2025.
7. G. Cao, J. Cao, and L. Xu, *Asymmetric multifractal scaling behavior in the Chinese stock market: Based on asymmetric MF-DFA*, Physica A: Statistical Mechanics and its Applications, vol. 392, no. 4, pp. 797–807, 2013.
8. L. Calvet and A. Fisher, *Multifractality in asset returns: Theory and evidence*, Review of Economics and Statistics, vol. 84, no. 3, pp. 381–406, 2002.
9. W. Chen, W. Hussain, F. Cauteruccio and X. Zhang, *Deep Learning for Financial Time Series Prediction: A State-of-the-Art Review of Standalone and Hybrid Models*, CMES-Computer Modeling in Engineering and Sciences, 2023.
10. P. Cho and M. Lee, *Forecasting the volatility of the stock index with deep learning using asymmetric Hurst exponents*, Fractal and Fractional, vol. 6, no. 7, pp. 394, 2022.
11. R. Cont, *Empirical properties of asset returns: stylized facts and statistical issues*, Quantitative Finance, vol. 1, no. 2, pp. 223, 2001.
12. R. F. Engle, *Autoregressive conditional heteroscedasticity with estimates of the variance of United Kingdom inflation*, Econometrica: Journal of the Econometric Society, pp. 987–1007, 1982.
13. T. Fischer and C. Krauss, *Deep learning with long short-term memory networks for financial market predictions*, European Journal of Operational Research, vol. 270, no. 2, pp. 654–669, 2018.
14. J. B. Florindo, R. R. Lima, F. A. dos Santos, and J. L. Alves, *GHENet: Attention-based Hurst exponents for the forecasting of stock market indexes*, Physica A: Statistical Mechanics and its Applications, vol. 667, p. 130540, June 2025.
15. C. Francq and J.-M. Zakoian, *GARCH models: structure, statistical inference, and financial applications*, John Wiley & Sons, 2019.
16. X.-L. Gao, Y.-H. Shao, Y.-H. Yang, and W.-X. Zhou, *Do the global grain spot markets exhibit multifractal nature?*, Chaos, Solitons & Fractals, vol. 164, pp. 112663, 2022.
17. Y. Hu, J. Ni, and L. Wen, *A hybrid deep learning approach by integrating LSTM-ANN networks with GARCH model for copper price volatility prediction*, Physica A: Statistical Mechanics and its Applications, vol. 557, 2020.
18. J. W. Kantelhardt, S. A. Zschiegner, E. Koscielny-Bunde, S. Havlin, A. Bunde, and H. E. Stanley, *Multifractal detrended fluctuation analysis of nonstationary time series*, Physica A: Statistical Mechanics and its Applications, vol. 316, no. 1-4, pp. 87–114, 2002.
19. W. Kristjanpoller, *Double asymmetric multifractal detrended fluctuation analysis (DA-MFDFA): Extending asymmetric multifractal methods for exogenous time series*, Physica A: Statistical Mechanics and its Applications, vol. 680, pp. 131027, 2025.
20. L. Kristoufek, *Fractal markets hypothesis and the global financial crisis: Scaling, investment horizons and liquidity*, Advances in Complex Systems, vol. 15, no. 06, pp. 1250065, 2012.
21. B. Mandelbrot, A. Fisher, and L. Calvet, *A multifractal model of asset returns*, 1997.
22. J. Michańków, Ł. Kwiatkowski, and J. Morajda, *Combining Deep Learning and GARCH Models for Financial Volatility and Risk Forecasting*, arXiv preprint arXiv:2310.01063, 2023.
23. A. K. Mishra, J. Renganathan, and A. Gupta, *Volatility forecasting and assessing risk of financial markets using multi-transformer neural network based architecture*, Engineering Applications of Artificial Intelligence, vol. 133, p. 108223, 2024.
24. N. Naifar, *Do global risk factors and macroeconomic conditions affect global Islamic index dynamics? A quantile regression approach*, The Quarterly Review of Economics and Finance, vol. 61, pp. 29–39, 2016.
25. A. B. Nasr, T. Lux, A. N. Ajmi, and R. Gupta, *Forecasting the volatility of the Dow Jones Islamic Stock Market Index: Long memory vs. regime switching*, International Review of Economics & Finance, vol. 45, pp. 559–571, 2016.
26. D. B. Nelson, *Conditional heteroscedasticity in asset returns: A new approach*, Econometrica, vol. 59, no. 2, pp. 347–370, 1991.
27. J. Patel, S. Shah, P. Thakkar, and K. Kotecha, *Predicting stock and stock price index movement using trend deterministic data preparation and machine learning techniques*, Expert Systems with Applications, vol. 42, no. 1, pp. 259–268, 2015.
28. C.-K. Peng, S. V. Buldyrev, S. Havlin, M. Simons, H. E. Stanley, and A. L. Goldberger, *Mosaic organization of DNA nucleotides*, Physical Review E, vol. 49, no. 2, pp. 1685, 1994.

29. V. Pipiras and M. S. Taqqu, *Long-range dependence and self-similarity*, vol. 45, Cambridge University Press, 2017.
30. E. Ramos-Pérez, P. J. Alonso-González, and J. J. Núñez-Velázquez, *Multi-Transformer: A new neural network-based architecture for forecasting S&P volatility*, Mathematics, vol. 9, no. 15, p. 1794, 2021.
31. F. Saâdaoui, *Testing for multifractality of Islamic stock markets*, Physica A: Statistical Mechanics and its Applications, vol. 496, pp. 263–273, 2018.
32. F. Saâdaoui, *Segmented multifractal detrended fluctuation analysis for assessing inefficiency in North African stock markets*, Chaos, Solitons & Fractals, vol. 181, pp. 114652, 2024.
33. J. Song, H. Lee, J. Lee, and W. Chang, *Forecasting realized volatility using deep learning quantile function*, Applied Soft Computing, vol. 175, 2025.
34. S. Drożdż, R. Gębarowski, L. Minati, P. Oświęcimka, and M. Wątorek, *Bitcoin market route to maturity? Evidence from return fluctuations, temporal correlations and multiscaling effects*, Chaos: An Interdisciplinary Journal of Nonlinear Science, vol. 28, no. 7, p. 071101, July 2018.
35. J. Sun, L. Wang and D. Gong, *An Adaptive Selection Method for Shape Parameters in MQ-RBF Interpolation for Two-Dimensional Scattered Data and Its Application to Integral Equation Solving*, Fractal and Fractional, vol. 7, no. 6, pp. 448, 2023.
36. Z. Tao, W. Wu, and J. Wang, *Series decomposition Transformer with period-correlation for stock market index prediction*, Expert Systems with Applications, vol. 237, p. 121424, 2024.
37. C. Tudor, *Non-Gaussian selfsimilar stochastic processes*, Springer, 2023.
38. L. Wang and R. S. T. Lee, *Stock Index Return Volatility Forecast via Excitatory and Inhibitory Neuronal Synapse Unit with Modified MF-ADCCA*, Fractal and Fractional, vol. 7, no. 4, pp. 292, 2023.
39. L. Yang, Y. Zhu, and Y. Wang, *Multifractal characterization of energy stocks in China: A multifractal detrended fluctuation analysis*, Physica A: Statistical Mechanics and its Applications, vol. 451, pp. 357–365, 2016.
40. X. Yu, D. Zhang, T. Zhu, and X. Jiang, *Novel hybrid multi-head self-attention and multifractal algorithm for non-stationary time series prediction*, Information Sciences, vol. 613, pp. 541–555, 2022.
41. Y. Yuan and T. Zhang, *Forecasting stock market in high and low volatility periods: A modified multifractal volatility approach*, Chaos, Solitons & Fractals, vol. 140, pp. 110252, 2020.
42. G. P. Zhang, *Time series forecasting using a hybrid ARIMA and neural network model*, Neurocomputing, vol. 50, pp. 159–175, 2003.
43. L. Zhang, F. Wang, B. Xu, W. Chi, Q. Wang, and T. Sun, *Prediction of stock prices based on LM-BP neural network and the estimation of overfitting point by RDCI*, Neural Computing and Applications, vol. 30, no. 5, pp. 1425–1444, 2018.