Article

Modeling and Forecasting Historical Volatility Using Econometric and Deep Learning Approaches: Evidence from the Moroccan and Bahraini Stock Markets

Imane Boudri * and Abdelhamid El Bouhadi

National School of Business and Management, Sidi Mohamed Ben Abdellah University, Fez 30000, Morocco; el_bouhadihamid@yahoo.fr
* Correspondence: imane.boudri@usmba.ac.ma

Abstract: This study challenges the prevailing belief in the necessity of complex models for accurate forecasting by demonstrating the effectiveness of parsimonious econometric models, namely ARCH(1) and GARCH(1,1), over deep learning robust approaches, such as LSTM and 1D-CNN neural networks, in modeling historical volatility within pre-emerging stock markets, specifically the Moroccan and Bahraini stock markets. The findings suggest reevaluating the balance between model complexity and predictive accuracy. Future research directions include investigating the potential existence of threshold effects in market capitalization for optimal model performance. This research contributes to a deeper understanding of volatility dynamics and enhances forecasting models’ effectiveness in diverse market conditions.

Keywords: historical volatility; pre-emerging markets; ARCH-GARCH models; deep learning approaches; LSTM network; 1D-CNN network

1. Introduction

In today’s rapidly evolving financial landscape, accurate forecasting of market volatility is crucial for investors and policymakers alike. Volatility, according to Granger and Poon (2001), denotes the degree of variability or dispersion of returns for a given security or market index over a specific period, while market volatility, according to Narula (2022), refers to the degree of variation in trading prices over time. Therefore, historical volatility serves as a gauge of the magnitude of past movements in a financial market and provides valuable insights into the market’s behavior over time. Understanding market volatility allows investors to predict patterns and future market movements and potentially earn profits (Burtniak and Suduk 2022; Li 2021). Predicting market volatility allows policymakers to effectively utilize fiscal and monetary policies to reduce systematic risk and then enhance their ability to stabilize financial markets (Yu et al. 2018).

The study of market volatility is vital due to its profound impact on financial decision-making. High volatility often signals increased risk and uncertainty, affecting investment strategies, portfolio management, and risk assessment. Conversely, low volatility may indicate stability but could also reflect a lack of market movement, influencing liquidity and trading activity. By accurately modeling and forecasting volatility, stakeholders can better manage these risks, optimize their investment portfolios, and make informed decisions that enhance market efficiency and stability.

Given the importance of precise volatility modeling and forecasting, it is imperative to explore and evaluate different modeling approaches. Traditional econometric models, such as ARCH and GARCH, have been the cornerstone of volatility analysis and forecasting (Bhowmik and Wang 2020). However, the advent of advanced machine learning and deep learning techniques has opened new avenues for improving forecast accuracy.
This study aims to contribute to the understanding of deep learning models in volatility modeling and forecasting by comparing their performance with traditional econometric models. Specifically, we examine historical volatility in the Moroccan and Bahraini stock markets, offering a comparative analysis of these methodologies. The inclusion of both markets allows for an assessment of potential differences in market structures, regulatory landscapes, and investor behaviors, providing a more comprehensive understanding of volatility dynamics in these emerging economies.

To our knowledge, this study is the first to model and forecast historical volatility in the context of pre-emerging markets in the MENA region using both econometric and machine learning approaches. Conventionally, researchers have shown a greater interest in modeling and forecasting returns, often prioritizing this over volatility. With the advent of artificial neural networks, research on forecasting returns has increased significantly, primarily because investors are more interested in returns. Despite this, understanding and predicting market volatility remains crucial. Volatility provides insights into the risk and stability of financial markets, which are essential for effective risk management and decision-making. In this study, we aim to address this gap by presenting a comprehensive review of the literature on volatility modeling, highlighting key econometric and machine learning techniques commonly employed in financial forecasting. This is followed by an overview of the Moroccan and Bahraini markets, emphasizing their unique characteristics, regulatory frameworks, and microstructure.

Our methodology section details the data sources, model specifications, and evaluation criteria used for performance assessment. We conduct an empirical analysis of historical volatility patterns in the Moroccan and Bahraini stock markets, comparing the forecasting precision of ARCH and GARCH models with that of LSTM and CNN architectures. Through this comparative analysis, we aim to identify the strengths and limitations of each approach, providing actionable insights for market practitioners and policymakers.

The findings of this study enrich the existing literature on volatility forecasting by providing empirical evidence on the performance of econometric and deep learning models in pre-emerging markets. Our research has significant implications for portfolio management, risk assessment, and financial decision-making in both academic and practical domains.

2. Literature Review

Studying the volatility of an asset or a financial market stands as one of the central pillars of modern financial research (Bhowmik and Wang 2020). This emphasis has only intensified since the market crash of 1987, which underscored the critical importance of understanding and effectively managing market volatility (Schwert 1990). As investors and policymakers grappled with the aftermath of this significant event, there was a heightened awareness of the profound impact that fluctuations in asset prices can have on financial stability, economic growth, and investor confidence. Consequently, researchers have increasingly focused on developing sophisticated models and analytical tools to better comprehend and predict market volatility, aiming to provide insights that can inform more robust risk management strategies and investment decisions.

Since the publication of the seminal paper of Engle (1982) introducing a new class of stochastic processes called autoregressive conditional heteroscedastic (ARCH) processes to estimate the variance of United Kingdom inflation, various studies have explored the effectiveness of different ARCH and GARCH family models in estimating volatility within stock markets. These models have demonstrated their efficacy in capturing the dynamics of asset volatility and also aggregate stock market volatility, due to their ability to capture the time-varying nature and clustering of volatility in financial time series data (Engle et al. 2013). Additionally, researchers have proposed distributed models to better characterize the thick tail of daily return rates. Engle (1982) initially introduced the autoregressive conditional heteroscedasticity model (ARCH model) to account for possible
correlations in the conditional variance of prediction errors, later extended by Bollerslev (1986) into the generalized autoregressive conditional heteroskedastic model (GARCH model). Subsequently, the GARCH model evolved into a family of models, including both linear (symmetric) and nonlinear (asymmetric) variants (such as EGARCH, TGARCH, or APARCH) to accommodate various characteristics of volatility.

Moreover, the GARCH framework accommodates the phenomenon of volatility persistence, where past volatility levels influence future volatility, making it particularly suitable for modeling the dynamics of financial market data (Alqaralleh et al. 2020). Additionally, the flexibility of the GARCH model in capturing both short- and long-term volatility patterns makes it a versatile tool for researchers seeking to understand and predict market dynamics (Bhowmik and Wang 2020). Overall, the GARCH model’s ability to capture the complex dynamics of volatility, including its time variation and persistence, makes it a preferred choice for many researchers in financial econometrics.

In the realm of modeling and forecasting stock market volatility, particularly in emerging and pre-emerging markets, empirical investigations into different GARCH models have yielded contradicting outcomes. On the one hand, a plethora of studies have shown that nonlinear models, especially the EGARCH model, outperformed their linear counterparts in capturing and predicting the conditional variance across short- and long-term horizons (Lin 2018; Abdalla and Winker 2012; Liu and Hung 2010; Alberg et al. 2008; Selçuk 2005; Chong et al. 1999). However, Al Rahahleh and Kao (2018) evaluated the forecasting performance of six linear and nonlinear GARCH-class models considering the Tadawul All Share Index (TASI) and the Tadawul Industrial Petrochemical Industries Share Index (TIPISI) for petrochemical industries in the Saudi Stock Market, using daily price data from 10 September 2007 to 26 February 2015. Their findings indicate that the asymmetric power ARCH (APARCH) model is the most accurate for forecasting volatility in both indices.

By contrast, limited evidence in the literature supports the efficacy of linear or symmetric volatility models, especially within the context of emerging markets (Srinivasan and Ibrahim 2010; Gaio et al. 2014). However, the literature has prominently highlighted the effectiveness of the parsimonious linear GARCH(1,1) model with a generalized distribution of residual, particularly in capturing the intricate dynamics of volatility within financial markets. Numerous empirical studies have underscored the model’s robustness and predictive power, emphasizing its ability to effectively model the volatility clustering and the conditional variance of asset returns and time series data across various time horizons (Zabiulla 2015; Abdalla and Winker 2012; Joshi 2010; Gokcan 2000).

In line with this, Sharma et al. (2021) conducted a comprehensive study comparing linear and nonlinear GARCH models for forecasting volatility in major emerging countries, namely China, India, Indonesia, Brazil, and Mexico. Using data from January 2000 to December 2019 and incorporating structural breaks due to the 2008 financial crisis, their findings indicate the superiority of the GARCH(1,1) model for volatility forecasting, with insignificant leverage effects.

More recently, numerous studies have delved into the realm of modeling and forecasting market volatility through the lens of deep learning models. Moon and Kim (2019) proposed an accurate algorithm for forecasting stock market indices and their volatility using the long short-term memory (LSTM) deep learning algorithm. Their study examines data from five stock market indices—S&P 500, NASDAQ, German DAX, Korean KOSPI200, and Mexico IPC—over a seven-year period from 2010 to 2016. Their results indicate that the highest prediction performance is achieved using a hybrid momentum approach, which considers the difference between the price and the moving average of past prices, for both market index and volatility predictions. Notably, unlike stock indices, the prediction accuracy for volatility does not depend on other financial variables such as open, low, high prices, or volume, but solely on the volatility itself. Lin and Sun (2021) introduced the SP-M–Attention model, a sparsely optimized, multi-headed, self-attentive neural network designed for volatility prediction, demonstrating superior performance across various financial forecasting tasks and highlighting its potential applicability to diverse time series
problems beyond finance. Cho and Lee (2022) addressed volatility forecasting in the stock market using asymmetric fractality features with deep learning algorithms. Their study introduces a novel approach employing asymmetric Hurst exponents to capture long-range dependence behavior, showing superior predictive power during high volatility periods, such as the COVID-19 pandemic in 2020.

Comparative research has pitted deep learning-based forecasters, including Multi-layer perceptrons, recurrent neural networks, temporal convolutional networks, and the temporal fusion transformer, against traditional econometric models like GARCH ones (Ge et al. 2023). These investigations have frequently revealed the superior predictive performance of deep learning architectures in capturing intricate long-range dependencies, often surpassing the performance of classical approaches in volatility forecasting in financial markets (Sahiner et al. 2023). In this context, Jia and Yang (2021) utilized deep neural network (DNN) and long short-term memory (LSTM) models to forecast stock index volatility. They introduced a likelihood-based loss function to enhance forecasting accuracy, demonstrating that their LSTM model outperforms econometric models and other deep learning approaches using distance loss functions.

Moreover, the integration of deep learning algorithms such as LSTM and GRU with sentiment data has yielded remarkable advancements, showcasing substantial improvements over GARCH models (Yu et al. 2023). Furthermore, the amalgamation of optimized variational mode decomposition with deep learning frameworks like DBN, LSTM, and GRU has unveiled enhanced predictive capabilities applicable across both emerging and developed markets (Cai et al. 2023). These collective findings underscore the efficacy of deep learning models in adeptly modeling and forecasting market volatility, illuminating their pivotal role in this domain.

However, studies focusing on the prediction of volatility in emerging and pre-emerging markets using deep learning methods remain relatively modest. Researchers often concentrate on developed markets, which are known for their high volatility and availability of high-frequency data.

3. Overview of the Moroccan and Bahraini Stock Markets

3.1. Casablanca Stock Exchange (CSE)

Established in 1929, the Casablanca Stock Exchange (CSE) is the third oldest marketplace in Africa, playing a central role in the regional economy. Over the years, the exchange has undergone significant transformations to modernize its operations and align with international standards. In 1995, a management company was established, entrusted with overseeing the stock exchange, operating under specifications approved by the Ministry of Finance. This move marked a pivotal moment in the evolution of the CSE, setting the stage for enhanced governance and strategic management. The regulatory environment is overseen by the Moroccan Capital Market Authority (AMMC), which ensures compliance with market regulations and investor protection, further bolstering the exchange’s credibility and stability. Despite these advancements, the CSE continues to face challenges in terms of liquidity, market depth, and market activity, necessitating ongoing efforts to strengthen market infrastructure and improve investor confidence.

With 77 listed companies as of the end of 2023, the Casablanca Stock Exchange (CSE) operates as a continuous market, characterized by the continuous matching of buy and sell orders throughout the trading day. Additionally, the CSE features an opening and closing fixing, which serve as reference points for the market at the beginning and end of each trading session. Driven by orders and facilitated by an intricate order book system, the market ensures that buy and sell orders are executed efficiently, contributing to market transparency and fair pricing.

3.2. Bahrain Bourse (BHB)

Transitioning to the Arabian Peninsula and venturing into the Asian continent, we explore the Bahrain Bourse, a younger financial institution established in 1987, character-
ized by its unique blend of traditional Islamic finance principles and modern investment practices. Despite its relatively recent inception, the Bahrain Stock Exchange has rapidly emerged as a significant player in the region’s financial landscape, with 42 listed companies as of the end of 2023.

In the Bahraini equity market, the order-driven mechanism, married to continuous auctions, serves as the backbone of trading activities, fostering efficient price discovery. Market participants, including individual investors, institutional traders, and foreign entities, have direct access to the trading platform, enabling them to place orders according to their investment strategies and market outlook. This decentralized approach promotes market depth and liquidity, as the continuous matching of orders facilitates swift execution and minimal price impact.

While the Bahraini market primarily operates as an order-driven system, it also incorporates elements of a quote-driven market, especially in the presence of market makers. These entities provide liquidity by quoting bid and ask prices for certain securities, helping to narrow spreads and improve market efficiency. However, the order-driven nature remains predominant, emphasizing the importance of investor orders in driving market dynamics and price formation.

However, the Bahraini equity market faces certain difficulties. One challenge is the limited number of listed companies, which can restrict investment choices and diversification opportunities for investors. Furthermore, market liquidity can vary across different securities, impacting trade execution and price stability, especially for less liquid stocks.

In fine, both the Moroccan and Bahraini markets are classified as pre-emerging markets by MSCI (Morgan Stanley Capital International), indicating their status as evolving financial hubs with potential for growth and development. Despite facing challenges, they offer promising investment opportunities for those seeking exposure to emerging economies. As they continue to implement reforms and enhance infrastructure, these markets are positioned to attract greater investment inflows in the future.

4. Materials and Methods

In this study, we embarked on a comparative analysis of volatility modeling and forecasting techniques, drawing insights from both econometric and deep learning methodologies. To achieve this goal, our methodology consisted of employing a combination of traditional econometric methods, namely autoregressive conditional heteroscedasticity (ARCH) and generalized autoregressive conditional heteroscedasticity (GARCH) models, alongside cutting-edge deep learning architectures such as long short-term memory (LSTM) and convolutional neural network (CNN) models. The subsequent sections present and elaborate on the various approaches utilized in this study for volatility modeling and forecasting in the two markets.

4.1. Econometric Approach

4.1.1. The Autoregressive Conditional Heteroscedastic (ARCH) Model

ARCH models involve time series characterized by changing volatility over time (heteroscedasticity), which is conditional on the autocorrelation of previous lags (autoregressive conditional). Referring to the seminal paper by Engle (1982), a variable $Y_t$ follows an ARCH ($p$) process if

\[ y_t | \psi_{t-1} \sim N(0, h_t) \]  
\[ h_t = a_0 + \sum_{i=1}^{p} a_i y_{t-i}^2 \]

where $\psi_{t-1}$ is the information set available at time $t-1$, $h_t$ is the conditional variance function, $p$ is the order of the ARCH process, and $\alpha$ is a vector of unknown parameters.

To guarantee positive variance, the parameters must satisfy some conditions, such that $\forall t \in \mathbb{Z}, a_0 > 0$ and $a_i > 0$. Furthermore, the recent past is assumed to have more influence than older lags, so $a_1 > a_2 > \cdots > a_p$.

For estimation using our ARCH model, we followed the steps indicated in Figure 1.
where \( \psi_{t-1} \) is the information set available at time \( t-1 \) including the appropriate ARCH and GARCH components.

- **Step 3**: Checking for ARCH effect using a heteroscedasticity test.
- **Step 4**: Estimating the variance equation using an ARCH model.

**Figure 1.** Illustration of ARCH model estimation steps.

4.1.2. The Generalized Autoregressive Conditional Heteroscedastic (GARCH) Model

GARCH models were introduced in 1986 by Bollerslev as an extension of ARCH models. They incorporate both autoregressive and moving average terms in the volatility equation. In addition to past squared residuals, GARCH models also include lagged conditional variances in the model specification. This allows them to capture both short-term volatility clustering and long-term persistence in volatility, making them more flexible and capable of modeling various volatility patterns observed in financial time series (Francq and Zakoian 2019).

Moreover, GARCH models provide a parsimonious alternative to high-order ARCH models, which can become problematic when estimating many ARCH effects (Bhowmik and Wang 2020).

The formalities of GARCH models involve both terms \( p \) and \( q \), whereas in ARCH, you only have the \( p \) term. So, a variable \( Y_t \) follows a GARCH \((p,q)\) process if

\[
y_t|\psi_{t-1} \sim N(0, h_t) \quad (3)
\]

\[
h_t = \alpha_0 + \sum_{i=1}^{p} \alpha_i y_{t-i}^2 + \sum_{i=1}^{q} \beta_i h_{t-i} \quad (4)
\]

where \( \alpha_0 > 0, \alpha_i > 0 \) and \( \beta_i > 0 \).

For estimation using our GARCH model, we followed the steps shown in Figure 2.

- **Step 1**: Estimating the mean equation using an ARIMA model.
- **Step 2**: Checking for ARCH effect using a heteroscedasticity test.
- **Step 3**: Re-estimating the model including the appropriate ARCH and GARCH components.
- **Step 4**: Conducting diagnostic tests.

**Figure 2.** Illustration of GARCH model estimation steps.

In summary, the conditional variance equation in ARCH and GARCH models serves to capture the time-varying nature of volatility in financial market data. This specification is essential in financial contexts, where analysts seek to predict future volatility based on past observations and information. In these models, the conditional variance at each time point is estimated as a weighted average of several components (Abdalla and Winker 2012). Firstly, a long-term average (represented by the constant term) provides a baseline estimate of volatility. Secondly, the GARCH term incorporates information from the previous period’s forecast variance, contributing to the persistence of volatility over time. Finally,
the ARCH term captures the impact of unexpected asset returns on volatility, adjusting the estimate of variance based on the magnitude and direction of these returns. This dynamic interplay between past volatility, current information, and unexpected returns allows ARCH and GARCH models to effectively capture the complex dynamics of financial market volatility.

4.2. Deep Learning Approach

If in the econometric approach, volatility is modeled as the conditional variance using ARCH and GARCH models, in the neural network approach, volatility is approximated by the standard deviation of daily logarithmic returns (close-to-close) for both indices. This is a straightforward calculation and a widely recognized measure of volatility in finance, offering an intuitive interpretation as it directly represents the magnitude of typical price fluctuations. This approach not only provides a clear and intuitive measure of volatility but also facilitates a comparison of model performance across different methodologies. Consequently, these series will serve as inputs for the LSTM and CNN models.

4.2.1. LSTM Network

Modeling and predicting future values of time series using artificial neural networks presents a challenge in retaining past information for accurate forecasting. Recurrent neural networks (RNNs) address this challenge by incorporating memory mechanisms in their hidden layers, allowing them to store and recall past states. However, simple RNNs are prone to gradient explosion, where gradients grow uncontrollably during training, affecting model stability (Kanai et al. 2017). To overcome this limitation, long short-term memory (LSTM) networks were developed. LSTM networks feature gated mechanisms that regulate the flow of information, preventing gradient explosion and enabling effective long-term memory retention. These networks learn to prioritize relevant information for prediction, enhancing forecasting accuracy. LSTM networks have proven effective for financial time series forecasting due to their ability to handle sequential data and long-term dependencies (Yu et al. 2019). However, training an LSTM network requires careful parameter selection to optimize performance (Sako et al. 2022).

In our study, we developed and trained a neural network with a range of parameters, as depicted in Table 1.

<table>
<thead>
<tr>
<th>Network Parameters</th>
<th>Values</th>
</tr>
</thead>
<tbody>
<tr>
<td>Data standardization formula</td>
<td>$x' = (x - \mu) / \sigma$</td>
</tr>
<tr>
<td>Optimization algorithm</td>
<td>Adam</td>
</tr>
<tr>
<td>Number of iterations (epochs)</td>
<td>250</td>
</tr>
<tr>
<td>Gradient threshold</td>
<td>1</td>
</tr>
<tr>
<td>Initial learning rate</td>
<td>0.005</td>
</tr>
<tr>
<td>Learning rate drop period</td>
<td>125</td>
</tr>
<tr>
<td>Learning rate drop factor</td>
<td>0.2</td>
</tr>
<tr>
<td>Number of hidden layers</td>
<td>100</td>
</tr>
<tr>
<td>Training rate</td>
<td>0.9</td>
</tr>
<tr>
<td>Testing rate</td>
<td>0.1</td>
</tr>
</tbody>
</table>

We trained the LSTM model over 250 epochs, where each epoch represents one complete pass through the dataset during training. A gradient threshold of 1 was applied to stabilize training by limiting the magnitude of gradients during backpropagation. The optimization algorithm chosen was Adam, known for its efficiency in stochastic optimization tasks. We initialized training with a learning rate of 0.005 and applied a learning rate drop every 125 epochs with a factor of 0.2 to optimize convergence. The models were structured with 100 hidden layers to capture complex patterns in the data. For evaluation, we used a training rate of 90% and a testing rate of 10% to assess model performance on realized data.
4.2.2. 1D-CNN Network

A convolutional neural network (CNN) is a versatile machine learning algorithm applied extensively in diverse fields, such as image processing, speech recognition, and time series analysis. Originally tailored for image data, CNNs necessitate two-dimensional input. However, given that time series data typically unfold in one dimension, a specialized adaptation known as a one-dimensional CNN has emerged (Markova 2022). This model operates on a single sequence, potentially incorporating multiple convolutional layers and a pooling layer to distill key features. Subsequently, a fully connected dense layer interprets these features, facilitated by a flattening layer to reduce dimensionality. To enhance performance, we meticulously curated and refined parameters using our comprehensive training dataset, with the objective of minimizing prediction errors, as illustrated in Table 2.

Table 2. CNN network parameters.

<table>
<thead>
<tr>
<th>Network Parameters</th>
<th>Values</th>
</tr>
</thead>
<tbody>
<tr>
<td>Data standardization formula</td>
<td>( x' = \frac{x - x_{\min}}{x_{\max} - x_{\min}} )</td>
</tr>
<tr>
<td>Optimization algorithm</td>
<td>Adam</td>
</tr>
<tr>
<td>Activation function</td>
<td>ReLU</td>
</tr>
<tr>
<td>Number of iterations (epochs)</td>
<td>1000</td>
</tr>
<tr>
<td>Gradient threshold</td>
<td>ReLU</td>
</tr>
<tr>
<td>Number of filter units</td>
<td>64</td>
</tr>
<tr>
<td>Number of kernels</td>
<td>2</td>
</tr>
<tr>
<td>Batch size</td>
<td>32</td>
</tr>
<tr>
<td>Dense units</td>
<td>1</td>
</tr>
<tr>
<td>Training rate</td>
<td>0.9</td>
</tr>
<tr>
<td>Testing rate</td>
<td>0.1</td>
</tr>
</tbody>
</table>

We trained the CNN network over 1000 epochs to capture temporal patterns in financial time series data. We utilized the Adam optimizer for efficient stochastic optimization and employed ReLU (rectified linear unit) as the activation function in convolutional layers to introduce nonlinearity. A gradient threshold of ReLU was applied to manage gradients during backpropagation. The CNN architecture featured 64 filter units across 2 kernels to extract spatial features from the input data. We utilized a batch size of 32 for training efficiency and configured the model with 1 dense unit for final output. Training and testing rates were set at 90% and 10%, respectively, to evaluate model performance on distinct datasets.

4.3. Forecast Performance Metrics

The performance of all employed models was evaluated using various metrics, which assessed the disparities between predicted and actual values. The performance metrics were as follows:

- Root mean square error (RMSE):\[
RMSE = \sqrt{\frac{1}{n} \sum_{t=1}^{n} (\epsilon_t)^2}
\] (5)

- Mean absolute error (MAE):\[
MAE = \frac{1}{n} \sum_{t=1}^{n} |\epsilon_t|
\] (6)

- Mean absolute percentage error (MAPE):\[
MAPE = \left( \frac{1}{n} \sum_{t=1}^{n} \frac{|\epsilon_t|}{x_t} \right) \times 100
\] (7)
5. Data and Descriptive Statistics

For the implementation of our study and the deployment of the selected models, we constructed two series of daily returns related to the performance of the Moroccan All Shares Index (MASI) and the Bahrain All Shares Index (BAX) over a 5-year period from January 2019 to December 2023. The data pertaining to the closing values of both indices were obtained from the Investing.com platform, ensuring reliable and comprehensive data coverage for our analysis. From these data, the daily logarithmic returns were computed using the following formula:

\[ R_{it} = \ln\left(\frac{P_i}{P_{i-1}}\right) \]

where \( P_t \) is the closing price of the index \( i \) at time \( t \).

Understanding the data through graphical representation in Figure 3 and descriptive statistics in Table 3 serves as a crucial foundation before proceeding to modeling and forecasting, providing valuable insights into the underlying patterns and characteristics of the datasets.

The descriptive statistics reveal insightful characteristics of the daily returns for the BAX and MASI. Notably, while the average daily return for BAX stands at 0.000321, MASI exhibits a slightly lower average return of 5.39 \times 10^{-5}. Both distributions display negative skewness, implying a left-skewed distribution with longer tails on the left side. Additionally, the high kurtosis values suggest heavier tails and more peaked distributions compared to a normal distribution, indicating greater volatility and potential for extreme returns. Furthermore, the Jarque–Bera test results decisively reject the hypothesis of normality for both indices, emphasizing the non-normal nature of their return distributions. Overall, it is noteworthy that there are no significant differences between the statistical characteristics of the two series, a point further reinforced by observing the graphical representations of the two datasets.
Table 3. Descriptive statistics of the BAX and MASI return series.

<table>
<thead>
<tr>
<th>Statistics</th>
<th>Variables</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>BAX</td>
</tr>
<tr>
<td>Number of observations</td>
<td>1226</td>
</tr>
<tr>
<td>Mean</td>
<td>0.000321</td>
</tr>
<tr>
<td>Median</td>
<td>0.000391</td>
</tr>
<tr>
<td>Maximum</td>
<td>0.034233</td>
</tr>
<tr>
<td>Minimum</td>
<td>-0.060013</td>
</tr>
<tr>
<td>Std. Dev.</td>
<td>0.005505</td>
</tr>
<tr>
<td>Skewness</td>
<td>-1.636707</td>
</tr>
<tr>
<td>Kurtosis</td>
<td>21.89118</td>
</tr>
<tr>
<td>Jarque–Bera</td>
<td>18777.81</td>
</tr>
<tr>
<td>Probability</td>
<td>0.000000</td>
</tr>
<tr>
<td>Normality hypothesis</td>
<td>Rejected</td>
</tr>
</tbody>
</table>

6. Empirical Results and Discussion

6.1. Empirical Results

6.1.1. Testing for Stationarity

To test the stationarity of our two-return series, we employed the augmented Dickey–Fuller (ADF) unit root test. This test is commonly used to determine whether a time series is stationary or nonstationary by assessing the null hypothesis regarding the presence of a unit root in the data. The results of the test are summarized in the Table 4 below, providing insights into the stationarity properties of the BAX and MASI return series.

Table 4. Stationarity test results for the BAX and MASI return series.

<table>
<thead>
<tr>
<th>Variables</th>
<th>ADF Test</th>
<th>T-Statistics</th>
<th>p-Values</th>
<th>Hypothesis</th>
</tr>
</thead>
<tbody>
<tr>
<td>BAX</td>
<td>Intercept</td>
<td>-15.88660 *</td>
<td>0.0000</td>
<td>Null hypothesis rejected</td>
</tr>
<tr>
<td></td>
<td>Trend and intercept</td>
<td>-15.88105 *</td>
<td>0.0000</td>
<td>Null hypothesis rejected</td>
</tr>
<tr>
<td></td>
<td>None</td>
<td>-15.81436 *</td>
<td>0.0000</td>
<td>Null hypothesis rejected</td>
</tr>
<tr>
<td>MASI</td>
<td>Intercept</td>
<td>-21.16787 *</td>
<td>0.0000</td>
<td>Null hypothesis rejected</td>
</tr>
<tr>
<td></td>
<td>Trend and intercept</td>
<td>-21.15932 *</td>
<td>0.0000</td>
<td>Null hypothesis rejected</td>
</tr>
<tr>
<td></td>
<td>None</td>
<td>-21.17456 *</td>
<td>0.0000</td>
<td>Null hypothesis rejected</td>
</tr>
</tbody>
</table>

Notes: * indicates significance at the 5% level.

The augmented Dickey–Fuller (ADF) test results indicate that the null hypothesis for both the BAX and MASI return series is rejected across all specifications. This suggests that the return series for both indices are stationary at level, indicating that they exhibit stable mean and variance over time.

6.1.2. Mean Equation Specification

To model the mean equation of both the Moroccan and Bahraini stock indices, we began with an analysis of the correlogram and evaluated various ARMA specifications. This process led us to select the ARMA(1,1) model as the most suitable, effectively capturing both the autoregressive and moving average components present in the daily return series. Initial empirical tests and iterative refinements confirmed the adequacy of the ARMA(1,1) model in representing the underlying dynamics of our stock market data. This robust representation of the conditional mean is crucial for our subsequent volatility modeling using ARCH and GARCH approaches. By accurately modeling the conditional mean, we can effectively isolate and analyze the volatility clustering effects that are central to our study.
6.1.3. Testing for Heteroscedasticity

We conducted the heteroscedasticity test using the ARCH effect specification, under the null hypothesis of no existing ARCH effect in the residual of ARMA(1,1) equation up to 1 lag. The results for both series are summarized in Table 5.

Table 5. Heteroscedasticity test results for the BAX and MASI residual return series.

<table>
<thead>
<tr>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>BAX</td>
<td>182.6009 *</td>
<td>0.0000</td>
<td>159.1391 *</td>
<td>0.0000</td>
<td>Null hypothesis rejected</td>
</tr>
<tr>
<td>MASI</td>
<td>37.99301 *</td>
<td>0.0000</td>
<td>36.92633 *</td>
<td>0.0000</td>
<td>Null hypothesis rejected</td>
</tr>
</tbody>
</table>

Notes: * indicates significance at the 5% level.

The results of the ARCH test, as indicated by the F statistics and chi-square statistics, reveal that the null hypothesis of no ARCH effect is rejected for both the BAX and MASI residual series. This suggests a pattern of volatility clustering, where significant changes are often succeeded by further significant changes, regardless of their direction, while minor changes tend to be followed by additional minor changes (Joshi 2010). Such findings have significant implications for modeling and forecasting, as ignoring heteroscedasticity can lead to biased parameter estimates and inaccurate predictions.

6.1.4. Estimation Results

Table 6 summarizes the predictive performance of the various models employed in our analysis. The selection of the ARCH(1) and GARCH(1,1) models was based on their ability to capture time-varying volatility in our financial return series. In our analysis, we extended the lag length beyond the first lag, observing that additional lags did not contribute significantly to the ARCH effect. This observation was validated by conducting the ARCH-LM test statistics at various lag lengths. The statistically insignificant p-values obtained for lag lengths beyond the first lag suggest that there is no ARCH effect remaining in the models. This implies that the chosen ARCH(1) and GARCH(1,1) specifications adequately capture the dynamics of volatility in the BAX and MASI return series.

Table 6. Historical volatility prediction: comparative analysis for different models.

<table>
<thead>
<tr>
<th></th>
<th>ARCH(1)</th>
<th>GARCH(1,1)</th>
<th>LSTM</th>
<th>1D CNN</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>BAX</td>
<td>MASI</td>
<td>BAX</td>
<td>MASI</td>
</tr>
<tr>
<td>RMSE</td>
<td>0.0055</td>
<td>0.0081</td>
<td>0.0055</td>
<td>0.0081</td>
</tr>
<tr>
<td>MAE</td>
<td>0.0033</td>
<td>0.0049</td>
<td>0.0034</td>
<td>0.0049</td>
</tr>
<tr>
<td>MAPE (%)</td>
<td>154.913</td>
<td>166.035</td>
<td>161.52</td>
<td>157.854</td>
</tr>
</tbody>
</table>

Observing the results, ARCH(1) and GARCH(1,1) models perform better in terms of RMSE, MAE, and QLIKE, indicating that they provide more accurate volatility predictions for both BAX and MASI.

LSTM and 1D CNN models have higher errors (RMSE and MAE) and less negative QLIKE values, suggesting they are less suitable for this specific volatility forecasting task. Despite the high MAPE for ARCH and GARCH, their lower absolute errors and QLIKE values make them preferable over neural network models in this context.

These findings suggest that the parsimonious ARCH and GARCH models are more effective in capturing the volatility dynamics of the BAX and MASI compared to the more complex LSTM and CNN models. Despite their computational complexity, the deep learning models do not exhibit superior predictive performance in this context.
6.2. Discussion

At first glance, our results may appear striking. The outperformance of parsimonious econometric models over robust deep learning models may seem counterintuitive, especially given the complexity and computational power associated with deep learning algorithms.

These results deviate from findings in several studies, such as the study by Petrozziello et al. (2022), which demonstrated the superiority of LSTM networks over various traditional GARCH models in predicting volatility for the NASDAQ 100 index. Similarly, in line with this perspective, Liu (2019) demonstrated that LSTM networks can provide superior predictions for longer time intervals compared to GARCH models.

Our findings, however, contradict these studies, demonstrating the effectiveness of simplistic models in volatility modeling and prediction. Thus, we align with researchers who emphasize the particular performance of the linear GARCH(1,1) model in capturing financial market volatility dynamics (Tripathy and Garg 2013). In addition, we highlight the remarkable performance of the parsimonious ARCH(1) model, which further strengthens our argument for the efficacy of simplistic models in volatility analysis.

However, we emphasize that our results hold true within the specific context of pre-emerging markets, which are already characterized by a lack of dynamism and liquidity. Additionally, it is important to note that we did not encounter similar studies within this particular context to facilitate comparison.

In conclusion, our findings underscore the crucial balance between model complexity and predictive accuracy in volatility forecasting within financial time series analysis. While many researchers advocate for the unconditional superiority of artificial neural network models in financial time series modeling and prediction, there is a growing recognition of the need for a deeper understanding of data characteristics. Aminimehr et al. (2022) label this realization as “inconsistency” and argue that the success of neural networks over econometric methods often stems from improper model implementation. In fine, despite significant advancements in this field, it is essential to acknowledge the continued effectiveness of less complex traditional methods.

7. Conclusions and Perspectives

In conclusion, we emphasize that our study sheds light on the nuanced relationship between model complexity and predictive accuracy in historical volatility forecasting within the context of pre-emerging markets. Our results advocate reevaluating the need for complex models, suggesting that simpler models may offer more precise results, and challenging the prevailing assumption that increased complexity is always necessary for accurate forecasts.

Our study has significant implications for both academics and practitioners. It suggests that in pre-emerging markets, where data quality and quantity might be limited, simpler models such as GARCH(1,1) can outperform more complex neural network-based models. This challenges the conventional wisdom that increasing model complexity invariably enhances predictive accuracy. By demonstrating that simpler models can be more effective in certain contexts, we provide a compelling case for the strategic selection of forecasting models based on the specific characteristics of the market being analyzed.

Looking ahead, an intriguing avenue for future research involves investigating the potential existence of a threshold effect. Specifically, exploring whether there is an optimal market capitalization threshold beyond which neural network models might demonstrate superior forecasting quality. Given the established superiority of neural networks in predicting volatility, returns, and prices in more developed markets, as documented in the literature, exploring this threshold effect could provide valuable insights into the applicability of deep learning approaches in different market conditions. Such research could involve a comparative analysis of various market segments, ranging from small-cap to large-cap stocks, to identify the precise conditions under which the benefits of neural network models become apparent.
Moreover, future studies could extend our analysis by incorporating other emerging and pre-emerging markets, as well as different asset classes. This broader scope would help determine whether our findings are generalizable across various financial contexts or specific to the markets we studied. Additionally, incorporating high-frequency data, where available, could provide further insights into the dynamic nature of volatility and the performance of different models over shorter time horizons.

In summary, our findings underscore the ongoing debate surrounding the choice of model complexity in financial time series analysis in general and historical volatility in particular and highlight the need for further exploration into the interplay between market characteristics and predictive modeling techniques. By addressing these gaps, future research can contribute to a more comprehensive understanding of volatility dynamics and enhance the effectiveness of forecasting models in diverse market environments and conditions. Our study serves as a foundational step in this direction, encouraging continued investigation into the optimal balance between model simplicity and complexity in financial forecasting.

Author Contributions: Both authors contributed equally to the work, including conceptualization, methodology, data analysis, and writing of the manuscript. All authors have read and agreed to the published version of the manuscript.

Funding: This research received no external funding.

Data Availability Statement: The data used in this study are publicly available and accessible through the website https://www.investing.com/ (accessed on 14 March 2024).

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

References


Ma, Wenfeng, Yuxuan Hong, and Yuping Song. 2024. On Stock Volatility Forecasting under Mixed-Frequency Data Based on Hybrid RR-MIDAS and CNN-LSTM Models. *Mathematics* 12: 1538. [CrossRef]


Zhang, Chao, Xingyue Pu, Mihai Cucuringu, and Xiaowen Dong. 2023. Graph Neural Networks for Forecasting Multivariate Realized Volatility with Spillover Effects. *arXiv*. [CrossRef]

**Disclaimer/Publisher’s Note:** The statements, opinions and data contained in all publications are solely those of the individual author(s) and contributor(s) and not of MDPI and/or the editor(s). MDPI and/or the editor(s) disclaim responsibility for any injury to people or property resulting from any ideas, methods, instructions or products referred to in the content.