International Journal of Statistical Sciences *ISSN 1683-5603*
Vol. 25(2), November, 2025, pp 173-178
DOI: https://doi.org/10.3329/ijss.v25i2.85780

A Comparative Study of GARCH and Deep Learning Models in Predicting Bitcoin Daily Returns

Md. Ziaul Hassan* and Md. Amir Hamja

Department of Statistics, Hajee Mohammad Danesh Science and Technology University, Dinajpur, Bangladesh
Email: amirhamja1190@gmail.com
*Correspondence should be addressed to Md. Ziaul Hassan
(Email: nirstathstu@gmail.com)

[Received August 11, 2025; Accepted October 26, 2025]

Abstract

Bitcoin has rapidly emerged as a focal point for the media, investors, and researchers due to its prominent role as an investment alternative to traditional currencies. However, its marked price volatility presents notable risks, particularly for organizations with significant Bitcoin holdings. To manage these risks effectively and enhance trading insights, accurate forecasting of Bitcoin's price fluctuations is crucial. This study presents a comparative analysis of Generalized Autoregressive Conditional Heteroskedasticity (GARCH) and Deep Learning (DL) models in predicting daily Bitcoin returns, aiming to identify the most effective approach for this highly volatile asset. Three variations of GARCH models, Exponential GARCH (EGARCH), Threshold GARCH (TGARCH), and Glosten-Jagannathan-Runkle GARCH (GJR-GARCH) were utilized, each evaluated under three distributional assumptions: Normal, student t, and Skewed student t. Additionally, four DL models Long Short-Term Memory (LSTM), Bidirectional LSTM (BiLSTM), Multilayer Perceptron (MLP), and Gated Recurrent Unit (GRU) were implemented to assess the efficacy of neural networks in volatility and return prediction. Model performance was measured using Root Mean Square Error (RMSE) and Root Mean Square Percentage Error (RMSPE), The results show that EGARCH with t-distribution achieved the lowest RMSE (2.720574) among the GARCH models, while MLP had the best overall performance among deep learning models with an RMSE of 2.731826 and the lowest RMSPE (3.431030) across all models. These findings indicate that both GARCH and DL models offer valuable insights, with EGARCH and MLP excelling in different performance metrics, suggesting complementary benefits in predicting Bitcoin returns.

Keywords: Bitcoin, Daily Return, GARCH, Deep Learning.

AMS Classification: 62P20, 68T07.

1. Introduction

Bitcoin has garnered significant attention as a digital asset, recognized for its role as an alternative investment to traditional currencies. However, its price is characterized by extreme volatility, presenting substantial risk to investors and institutions holding large Bitcoin portfolios. Accurate prediction of Bitcoin's price fluctuations is therefore crucial for developing informed investment and risk management strategies (Katsiampa, 2017). Traditionally, statistical models such as the

Generalized Autoregressive Conditional Heteroskedasticity (GARCH) model and its variants including EGARCH, TGARCH, and GJR-GARCH have been widely employed to forecast financial market volatility. These models effectively capture key characteristics such as volatility clustering and asymmetric volatility commonly observed in asset markets (Engle, 1982; Nelson, 1991). In the context of Bitcoin, GARCH-type models have proven useful for analyzing volatility under various distributional assumptions, such as Normal and t-distributions, which account for the frequent and extreme price fluctuations inherent in cryptocurrency markets (Khedr et al., 2021).

Empirical research has reinforced the relevance of these econometric models. Katsiampa (2017) demonstrated that GARCH-type models, particularly EGARCH and TGARCH, effectively model Bitcoin's asymmetric volatility behavior. Similarly, Chu et al. (2017) conducted a comparative study among multiple GARCH variants and concluded that complex structures, such as GJR-GARCH, perform better in capturing high-stress market conditions. Despite these advancements, traditional econometric models remain constrained by their linear assumptions and limited capacity to capture the nonlinear and chaotic dynamics that often characterize cryptocurrency markets.

To address these limitations, recent years have seen a surge in the use of machine learning (ML) and deep learning (DL) techniques for financial forecasting. ML approaches such as Random Forest, k-Nearest Neighbors (k-NN), and Extreme Gradient Boosting (XGBoost) have demonstrated strong predictive capabilities by effectively modeling nonlinear dependencies in large and noisy datasets. For instance, Mallqui and Fernandes (2019) compared ARIMA with ML algorithms and found that ML models achieved higher accuracy in Bitcoin price prediction. Likewise, Patel et al. (2021) employed XGBoost and ensemble learning methods, achieving superior multi-step forecasting performance compared to traditional econometric models.

Building upon these developments, deep learning models have become increasingly prominent in cryptocurrency forecasting research. Fischer and Krauss (2018) and Huang et al. (2019) highlighted the potential of neural architectures such as Long Short-Term Memory (LSTM), Bidirectional LSTM (BiLSTM), Multilayer Perceptron (MLP), and Gated Recurrent Unit (GRU) networks for capturing temporal dependencies and complex patterns in volatile time-series data. Torres et al. (2021) showed that LSTM networks significantly outperform ARIMA and simple neural networks in forecasting Bitcoin price trends. Sebastião et al. (2021) extended this line of work by demonstrating that BiLSTM models, through their ability to process information bidirectionally, achieve higher predictive accuracy, while GRU models offer comparable results with lower computational cost, making them ideal for real-time financial applications.

More recently, hybrid modeling frameworks that integrate statistical and DL techniques have emerged as promising tools for cryptocurrency forecasting. These hybrid models combine the interpretability of econometric approaches with the pattern recognition and generalization strengths of deep learning. For example, Zahid et al. (2022) introduced an EGARCH-GRU hybrid model that successfully improved forecast accuracy by leveraging GRU's nonlinear learning ability with EGARCH's capacity for volatility asymmetry modeling. Similarly, Qiu et al. (2025) proposed a GARCH-LSTM hybrid model, achieving lower Root Mean Squared Error (RMSE) and Mean Absolute Error (MAE) values than standalone models. Such hybrid approaches not only enhance predictive precision but also provide practical insights into market dynamics, offering valuable tools for investors navigating the turbulent cryptocurrency landscape.

Against this backdrop, the present study conducts a comprehensive comparison between GARCH-based models (GARCH, EGARCH, TGARCH, and GJR-GARCH) and deep learning models (LSTM, BiLSTM, MLP, and GRU) to identify the most effective method for forecasting daily

Bitcoin returns. Using performance metrics such as Root Mean Squared Error (RMSE) and Root Mean Squared Percentage Error (RMSPE), the study reveals that the MLP model achieves the highest predictive accuracy. These findings contribute to the growing literature on cryptocurrency forecasting by underscoring the superior performance of deep learning techniques in capturing the complex and volatile nature of Bitcoin price movements, thereby offering valuable implications for investors, traders, and policymakers. The remaining of the article is structured as follows: Section 2 outlines the theoretical framework and research methodology. Section 3 presents and interprets the results. Lastly, Section 4 provides the conclusions along with relevant policy implications.

2. Theoretical framework and research methodology

This section outlines the tools and techniques applied in the study. First, we present an overview to illustrate the workflow, followed by a detailed explanation of each individual component.

2.1 Approach overview

Initially, daily Bitcoin price data is collected and then transformed into daily returns for preprocessing. Next, statistical tests and metrics are applied to assess the suitability of GARCH family models. Afterward, the data is split into estimation and forecasting sets, and various GARCH models with three types of standardized residual distributions, along with four different configurations of DL models, are evaluated. In the final step, the top-performing GARCH and DL models are compared for predictive accuracy. The detailed diagram of the proposed methodology is shown in Figure 1.

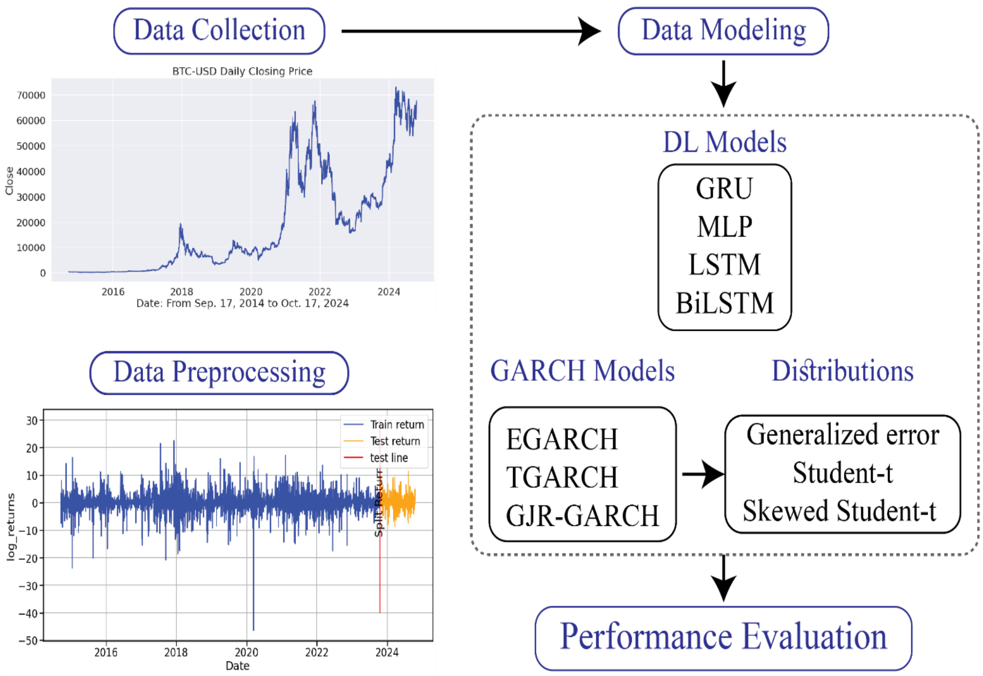

Figure 1: Detailed diagram of the proposed methodology in predicting bitcoin daily return.

2.2 Dataset Descriptions and Preprocessing

The dataset of daily Bitcoin price index in USD was obtained online from http://www.coindesk.com/price, covering the period from September 17, 2014, to October 17, 2024. The time series of daily closing prices, denoted as p_t, was converted into daily returns using the following equation.

$$r_t = \log\left(\frac{p_t}{p_{t-1}}\right) \tag{1}$$

The conditional mean of daily returns r_t follows an ARMA (p, q) model, as shown in equation below:

$$r_t = \mu + \sum_{i=1}^{p} a_i r_{t-i} + \sum_{j=1}^{q} b_j \epsilon_{t-j} + \epsilon_t \tag{2}$$

where the residuals ϵ_t form a time series with a non-constant variance, denoted by σ_t^2. The BTC returns dataset was split into 3314 observations for training (model estimation) and 368 for testing (forecast evaluation) across both parametric and nonparametric models.

2.3 Description of GARCH Family Models

In this study, we focus on volatility modelling using GARCH models. Engle introduced the ARCH model to capture the changing (conditional) variance in the error component, ε_t of a time series. Bollerslev then extended this to the GARCH model, where the residuals' conditional variance is modelled as,

$$\sigma_t^2 = \omega + \sum_{i=1}^{m} \alpha_i \epsilon_{t-i}^2 + \sum_{j=1}^{n} \beta_j \sigma_{t-j}^2 \tag{3}$$

where ω is a constant, α_i represents the ARCH terms capturing past squared residuals, and β_j represents the GARCH terms capturing past variances. To enhance the original GARCH model's flexibility and address asymmetry in response to shocks, a family of asymmetric models was introduced. These models assume that variance depends on both the sign and magnitude of shocks, with differences among them focusing on how asymmetries are modelled. Nelson proposed the EGARCH (exponential GARCH) model. Zakoian proposed another model called TGARCH (threshold GARCH). Similarly, Glosten proposed GJR-GARCH model which is the counterpart of TGARCH. In parametric modelling, we varied the distribution of the standardized residuals z_t by examining the impact of each of the following distributions:

- Generalized error distribution (ged): $z_t \sim GED(0, 1, \nu)$
- Student t-distribution (t): $z_t \sim t(0, 1, \nu)$
- Skewed Student t-distribution (skewt): $z_t \sim skt(0, 1, \nu, g)$

2.4 Description of DL Models

MLP is a fully connected feedforward neural network, useful for time series forecasting though limited in capturing sequential dependencies due to its fixed input structure. LSTM networks, on the other hand, are recurrent neural networks designed to capture long-term dependencies in sequential data, making them ideal for time series as they avoid vanishing gradient issues. BiLSTM extends the LSTM by processing data in both forward and backward directions, enhancing its ability to capture dependencies from both past and future contexts. The GRU, a simpler and more efficient alternative to LSTM, uses gating mechanisms for information control and memory retention, effectively capturing sequential patterns with fewer parameters.

2.5 Performance Measuring Techniques

For evaluating time series forecasting accuracy, we used RMSE and RMSPE. The former calculates the square root of the average squared differences between predicted and actual values, penalizing larger errors and emphasizing absolute accuracy. In contrast, RMSPE measures the average percentage error, providing a scaled metric that accounts for relative error sizes, making it useful for assessing accuracy across different levels of the forecasted variable.

3. Results and discussion

Table 1 displays the RMSE values for EGARCH, TGARCH, and GJR-GARCH when applied to different residual distributions for volatility forecasting. Each row corresponds to a specific distribution: t-distribution (t), Generalized Error Distribution (ged), and skewed t-distribution (skewt). Among the models, EGARCH with t-distribution has the lowest RMSE (2.720574), indicating relatively higher accuracy, while other model-distribution combinations yield slightly higher but comparable RMSE values, reflecting similar predictive performance across the configurations tested.

Table 1: Performance of the GARCH Models in terms of RMSE

Distribution	EGARCH	TGARCH	GJR-GARCH
t	2.720574	2.724133	2.723861
ged	2.724355	2.723841	2.723371
skewt	2.723717	2.724419	2.723828

Also, Table 2 presents the RMSPE values for GARCH family models. EGARCH with ged distribution has the lowest RMSPE (4.880723), suggesting superior performance among the configurations. The TGARCH and GJR-GARCH models generally show higher RMSPE values, especially under the t and skewt distributions, indicating relatively higher forecasting errors in these cases.

Table 2: Performance of the GARCH Models in terms of RMSPE

Distribution	EGARCH	TGARCH	GJR-GARCH
t	6.610538	11.707514	12.219845
ged	4.880723	5.970294	6.356839
skewt	6.736800	12.001043	12.220708

Table 3: Performance of the DL Models in terms of RMSE and RMSPE

Models	RMSE	RMSPE
MLP	2.731826	3.431030
LSTM	2.731274	5.420292
BiLSTM	2.728029	8.595744
GRU	2.730138	6.471126

Table 3 shows the RMSE and RMSPE values for DL models used in this study. BiLSTM has the lowest RMSE (2.728029), suggesting it performs best in terms of absolute error, though it has a higher RMSPE (8.595744) compared to others, indicating more relative error. Conversely, MLP shows a low RMSPE (3.431030), suggesting it maintains good relative accuracy, while GRU and LSTM show moderate performance across both metrics.

5. Conclusions

This study demonstrates the effectiveness of both GARCH models and DL approaches in forecasting Bitcoin returns, crucial for managing its inherent volatility. The EGARCH model, particularly with t-distribution, exhibited strong performance in RMSE, while the MLP model achieved the lowest RMSPE, indicating its relative accuracy. The findings suggest that a hybrid approach, integrating GARCH for volatility assessment and MLP for capturing complex patterns, could enhance forecasting accuracy. This research contributes to cryptocurrency forecasting and highlights the value of robust models in informing trading strategies. Future studies may explore ensemble methods to combine the strengths of both modelling approaches for improved predictive performance.

Acknowledgement: With appreciation, the author acknowledges the reviewer's thoughtful review and kind support.

References

[1] Chu, J., Chan, S., Nadarajah, S. and Osterrieder, J. (2017). GARCH modelling of cryptocurrencies. Journal of Risk and Financial Management, 10(4):17.
[2] Engle, R. F. (1982). Autoregressive conditional heteroscedasticity with estimates of the variance of United Kingdom inflation. Econometrica: Journal of the econometric society, 987-1007.
[3] Fischer, T. and Krauss, C. (2018). Deep learning with long short-term memory networks for financial market predictions. European journal of operational research, 270(2):654-669.
[4] Huang, W., Nakamori, Y. and Wang, S. Y. (2005). Forecasting stock market movement direction with support vector machine. Computers & operations research, 32(10):2513-2522.
[5] Katsiampa, P. (2017). Volatility estimation for Bitcoin: A comparison of GARCH models. Economics letters, 158:3-6.
[6] Khedr, A. M., Arif, I., El-Bannany, M., Alhashmi, S. M. and Sreedharan, M. (2021). Cryptocurrency price prediction using traditional statistical and machine-learning techniques: A survey. Intelligent Systems in Accounting, Finance and Management, 28(1):3-34.
[7] Mallqui, D. C. and Fernandes, R. A. (2019). Predicting the direction, maximum, minimum and closing prices of daily Bitcoin exchange rate using machine learning techniques. Applied Soft Computing, 75:596-606.
[8] Patel, J., Patel, A., Thakkar, R. and Doshi, P. (2021). Forecasting of Bitcoin price using ensemble learning. Procedia Computer Science, 192:693-702.
[9] Qiu, Z., Kownatzki, C., Scalzo, F. and Cha, E. S. (2025). Historical Perspectives in Volatility Forecasting Methods with Machine Learning. Risks, 13(5):98.
[10] Sebastião, H. and Godinho, P. (2021). Forecasting and trading cryptocurrencies with machine learning under changing market conditions. Financial Innovation, 7(1):3.
[11] Torres, J. F., Hadjout, D., Sebaa, A., Martínez-Álvarez, F. and Troncoso, A. (2021). Deep learning for time series forecasting: a survey. Big data, 9(1):3-21.
[12] Zahid, M., Iqbal, F. and Koutmos, D. (2022). Forecasting Bitcoin volatility using hybrid GARCH models with machine learning. Risks, 10(12):237.