Zero : Jurnal Sains, Matematika, dan Terapan
E-ISSN : 2580-5754; P-ISSN : 2580-569X
Volume 10, Number 2, 2026
DOI: 10.30829/zero.v10i2.30726
Page: 731-742

Comparing Volatility and Neural Network Models for Forecasting Bitcoin Prices in Indonesian Rupiah

[1]Raihan Akbar
Applied Mathematics, IPB University, Bogor Regency, Indonesia

[2]I Wayan Mangku
Mathematics, IPB University, Bogor Regency, Indonesia

[3]Bib Paruhum Silalahi
Mathematics, IPB University, Bogor Regency, Indonesia

Article Info

Article history:

Accepted: 10 July 2026

Keywords:

Cryptocurrency;
eGARCH-X;
Hybrid model;
NARX;
Volatility forecasting.

ABSTRACT

As Indonesia emerges among Southeast Asia's largest cryptocurrency markets, forecasting Bitcoin priced in rupiah (BTC/IDR) has gained practical relevance, yet most research focuses on BTC/USD and overlooks domestic macroeconomic conditions. This study tests whether augmenting an Exponential GARCH with Exogenous Variables (eGARCH-X) model with a Nonlinear Autoregressive network with eXogenous inputs (NARX) improves forecasts of weekly BTC/IDR log returns, using IHSG, USD/IDR, gold price, and BI rate as exogenous inputs. Using 437 weekly observations from January 2018 to June 2026, the hybrid was benchmarked against eGARCH-X, NARX, and a restricted eGARCH without exogenous terms across three splits, evaluated with RMSE, MAE, directional accuracy, and the Diebold−Mariano test, with each network comparison replicated over ten random initializations. The eGARCH identified a well-determined conditional variance process: volatility was highly persistent (0.976), responded asymmetrically to the sign of innovations (0.038, $p = 0.010$), and displayed heavy tails, with standardized residuals passing all diagnostics. No model differed significantly from the eGARCH-X benchmark, directional accuracy was indistinguishable from chance throughout (43.7–55.5%), and the exogenous regressors were insignificant both inSeaksample and out-of-sample. Weekly BTC/IDR returns thus appear tractable in their variance but close to unforecastable in their conditional mean, locating the practical value of these models in volatility estimation rather than directional prediction.

Corresponding Author:

Raihan Akbar,
Applied Mathematics
IPB University
Email: akbarraihan@apps.ipb.ac.id

1. INTRODUCTION

Originally conceived as a peer-to-peer payment method, Bitcoin has become the most actively traded crypto-asset, increasingly used in portfolio allocation and hedging. Lacking a central authority, its value is driven largely by market demand, making prices highly sensitive to speculation, macroeconomic conditions, and investor sentiment [1], [2], the literature further reports weak-form inefficiency, pronounced volatility clustering, and contested evidence on whether Bitcoin behaves like gold and whether its volatility reacts asymmetrically to shocks [3], [4]. These features make accurate forecasting difficult yet valuable [5]. Most evidence concerns BTC/USD, however, while Indonesia one of Southeast Asia's fastest-growing crypto markets, with nearly 22 million users and trading volume exceeding Rp 650 trillion (about US$40 billion) in 2024 remains understudied [6]. This calls for forecasting methods suited to Bitcoin's volatility behavior.

Three methodological strands dominate forecasting: interpretable but linear models such as ARIMA/SARIMA [7]; the ARCH/GARCH family, whose eGARCH variant captures the leverage effect and who's exogenous ("-X") extensions admit macroeconomic regressors [8], [9], [10], [11]; and machine-learning models such as NARX, which exploit lagged outputs and exogenous inputs to capture nonlinear dependence [12], [13]. Whether combining these approaches actually improves forecasting, however, remains unsettled. This progress has not extended to BTC/IDR: GARCH models handle changing variance but not strong nonlinearity, few studies integrate eGARCH-X with NARX in a single framework, and prior work rarely accounts for Indonesian macroeconomic variables.

This study addresses that gap by testing rather than assuming whether augmenting an eGARCH-X model with a NARX component, incorporating Indonesian macroeconomic variables, improves out-of-sample forecast accuracy for weekly BTC/IDR returns. Let $e_t$denote the one-step-ahead forecast error and $L(\cdot)$the loss function, and let $d_t = L\left(e_t^{\text{eGARCH-X}}\right) - L\left(e_t^{\text{challenger}}\right)$denote the loss differential between the eGARCH-X benchmark and a challenger model at time t. The null hypothesis is one of equal predictive accuracy, $H_0: E[d_t] = 0$, against the one-sided alternative that the challenger is more accurate, $H_1: E[d_t] > 0$; a negative Diebold Mariano statistic therefore indicates that the challenger is less accurate than the benchmark. $H_0$is evaluated under squared-error and absolute-error loss across three train/test splits. Directional accuracy is compared as a secondary, descriptive criterion and is reported without a formal test.

The contribution is evaluative rather than architectural. First, it identifies a well-specified conditional variance process for weekly BTC/IDR persistent, asymmetric, and heavy-tailed and shows that this, rather than the conditional mean, is where the series' modellable structure lies. Second, it establishes a measured bound on mean-predictability: three distinct extensions of the conditional mean, tested in-sample and out-of-sample, yield no statistically detectable gain, and directional accuracy does not differ from chance in any split. Third, it provides the first evidence that Indonesian macro-financial variables do not carry information for BTC/IDR a result of interest given the size of the domestic market. Fourth, it demonstrates that predictive-accuracy tests for stochastically trained models must be replicated across initializations: a single trained network in our own results produced a significant Diebold Mariano statistic that held in only three of ten replications.

2. RESEARCH METHOD

2.1 Data and Sources

Weekly rather than daily observations are used to reduce short-term noise while retaining economically meaningful variation, consistent with common practice in cryptocurrency forecasting studies [7], [14]. The outcome variable is the weekly BTC/IDR log return, modeled as a function of four Indonesian macroeconomic predictors: the IHSG index (domestic equity performance and investor confidence), the USD/IDR exchange rate (external monetary conditions and currency pressure), the domestic gold price (a competing safe-haven asset), and the Bank Indonesia interest rate. These variables capture distinct dimensions of the domestic financial environment that the literature links to cryptocurrency dynamics [15], [16], [17], [18], [19].

All market price series were retrieved from Yahoo Finance, with two exceptions detailed below. The BTC/IDR series was constructed as the product of the daily BTC/USD closing price and the contemporaneous USD/IDR exchange rate, with both series matched on calendar date prior to aggregation to prevent temporal misalignment in the cross-rate computation. Gold prices were retained in USD terms rather than converted to IDR, since the USD/IDR rate already enters the model as a separate regressor; converting gold to IDR would mechanically embed the exchange rate's movement into both series and induce spurious correlation between them. The Bank Indonesia policy rate, revised at irregular intervals by the Board of Governors, was obtained from Bank Indonesia's official statistical database. Daily series were aggregated to weekly frequency by retaining the last available closing observation within each Friday-ending week. Combined with the forward-filled BI policy rate, this yielded 438 weekly level observations from 5 January 2018 to 26 June 2026 and, after first-differencing, a final panel of 437 weekly log returns from 12 January 2018 to 26 June 2026. Five calendar weeks within this span are absent from the panel because the IHSG recorded no trading days in those weeks all five coincide with the extended Eid al-Fitr holiday (15 June 2018, 7 June 2019, 6 May 2022, 12 April 2024, and 4 April 2025) and the inner join across series therefore excludes them. These weeks were not imputed: assigning an IHSG price while the exchange was closed would create an artificial return.

Before estimation, several preprocessing steps are applied. Each price series is transformed into logarithmic returns to stabilize variance and improve stationarity. The log return of a variable y_t is expressed in Equation (1).

$$r_t = \ln(y_t / y_{t-1}) \tag{1}$$

Table 1. Variables used in this study

Variable	Symbol	Source	Sample Period
Bitcoin Price (IDR)	BTC_t	Yahoo Finance (BTC/USD) × USD/IDR	5 January 2018 – 26 June 2026
IHSG	$IHSG_t$	Yahoo Finance ($\hat{}JKSE$)	5 January 2018 – 26 June 2026
USD/IDR Exchange Rate	FX_t	Yahoo Finance ($IDR = X$)	5 January 2018 – 26 June 2026
Gold Price (USD)	$GOLD_t$	Yahoo Finance ($GC = F$)	5 January 2018 – 26 June 2026
BI Rate	BI_t	Bank Indonesia	5 January 2018 – 26 June 2026

where r_t is the return at time t. Log returns are standard in financial econometrics because they reduce heteroskedasticity, facilitate inference, and yield more nearly stationary series than raw prices [7], [8]. Finally, each explanatory variable is min–max normalized before being supplied to the NARX network to improve numerical stability during training [20], [21].

2.2 eGARCH-X Model

Financial time series, and cryptocurrency prices in particular, exhibit volatility clustering, leptokurtosis, and time-varying conditional variance, which violate the constant-variance assumption of linear regression and standard autoregressive models [3]. To accommodate these features, conditional heteroskedastic models of the ARCH/GARCH family are used [8]. The standard GARCH specification, however, restricts the response of volatility to be symmetric in the sign of shocks, whereas Bitcoin returns frequently display asymmetric (leverage) effects in which negative and positive innovations affect conditional variance differently an effect for which the eGARCH model of Nelson is well suited [9], [22]. Evidence on the direction and significance of this asymmetry in cryptocurrency markets is mixed [4], [7], which further motivates an asymmetry-aware specification rather than a symmetric one.

Let r_t denote the logarithmic return of Bitcoin prices at time t. Equation (2) shows the conditional mean equation and Equation (3) defines the error term.

$$r_t = \mu + \sum_{i=1}^{p} \phi_i r_{t-i} + \sum_{j=1}^{q} \psi_j \varepsilon_{t-j} + \sum_{k=1}^{4} \gamma_k X_{k,t} + \varepsilon_t \tag{2}$$

where ϕ_i represents the AR coefficient, ψ_j represents the MA coefficient, γ_k is the coefficient on the k-th exogenous regressor in the conditional mean and ε_t denotes the error term satisfying

$$\varepsilon_t = \sigma_t z_t \tag{3}$$

with z_t an i.i.d. process of zero mean and unit variance and σ_t is the standard deviation at time t. To incorporate domestic macroeconomic information directly into the volatility process, Equation (4) specified as an eGARCH(1,1) model augmented with exogenous regressors (eGARCH-X) [9], [10]:

$$ln(\sigma_t^2) = \omega + \beta\, ln(\sigma_{t-1}^2) + \alpha(|z_{t-1}| - E|z_{t-1}|) + \theta z_{t-1} + \sum_{k=1}^{4} \delta_k X_{k,t}^2 \tag{4}$$

where $z_{t-1} = \varepsilon_{t-1}/\sigma_{t-1}$; ω is the constant; β measures volatility persistence; α governs the magnitude (size) effect of shocks; and θ is the asymmetry (leverage) parameter, with $\theta < 0$ indicating that negative shocks raise volatility more than positive shocks of equal size [8]. The logarithmic formulation guarantees $\sigma_t^2 > 0$ without imposing non-negativity constraints on the parameters. The exogenous variables are arranged into vector in Equation (5).

$$X_t = \begin{pmatrix} IHSG_t \\ FX_t \\ GOLD_t \\ BI_t \end{pmatrix} \tag{5}$$

and $\delta = (\delta_1, \delta_2, \delta_3, \delta_4)^T$ quantifies the contribution of each Indonesian macroeconomic variable to conditional volatility. Allowing the conditional variance to respond to domestic financial conditions rather than to past volatility alone yields a more flexible representation of Bitcoin market dynamics during macroeconomic stress [15], [23]. Parameters are estimated by Maximum Likelihood Estimation (MLE) using rugarch package in R assuming a skewed Student-t (sstd) conditional distribution for z_t to accommodate the excess kurtosis typical of cryptocurrency returns. The conditional mean follows $ARMA(p, q)$ where the lag orders $p, q \in \{0,1,2\}$ were selected jointly by minimizing the Akaike Information Criterion (AIC) over an exhaustive 3 × 3 grid estimated on the training set. In addition, the (1,1) lag specification follows standard practice in the eGARCH [9].

Although eGARCH-X captures conditional heteroskedasticity, asymmetry, and exogenous macroeconomic effects, the standardized residuals may retain nonlinear temporal dependencies that conventional econometric models do not represent [4], [5]. These residuals are therefore passed to a NARX network designed to learn such nonlinear structure and improve forecasting performance within the proposed hybrid framework.

2.3 Nonlinear Autoregressive Network with Exogenous Inputs (NARX)

To capture nonlinear dependencies left unexplained by eGARCH-X, this study employs a Nonlinear Autoregressive network with eXogenous inputs (NARX) [20]. NARX is a dynamic recurrent network that uses delayed values of both the target series and exogenous inputs, which makes it effective for nonlinear time-series forecasting and well established for financial and exchange-rate applications [21]. General form of NARX is expressed in Equation (6).

$$n_t = f\left(n_{t-1}, \dots, n_{t-d_y}, u_t, \dots, u_{t-d_u}\right) + a_t \quad (6)$$

where n_t is the (standardized residual) target series, $u_t = (IHSG_t, FX_t, GOLD_t, BI_t)$ is the exogenous input vector, d_y and d_u are the output and input delays, and $f(\cdot)$ is the nonlinear mapping approximated by the network.

The NARX architecture was optimized via an exhaustive grid search over the number of hidden neurons, $h \in \{5, 10, 20, 50\}$, and the input/feedback delay order, $d_y = d_u = d \in \{1, 2, 3, 4\}$. To mitigate the sensitivity of the Levenberg–Marquardt algorithm to weight initialization (maximum 1000 epochs), each of the sixteen (h, d) configurations was trained ten times with independent random initial weights, yielding 160 runs in total with mean squared error (MSE) validation. All inputs and targets were normalized to $[-1, 1]$ via mapminmax prior to training and denormalized after forecasting.

2.4 Hybrid eGARCH-X-NARX Framework

The hybrid forecast is assembled in three sequential steps. First, the fitted eGARCH-X model produces a one-step-ahead conditional mean forecast $\hat{\mu}_t$ and conditional standard deviation forecast $\hat{\sigma}_t$ for each out-of-sample period. Second, the in-sample standardized residuals $z_t = \frac{\varepsilon_t}{\hat{\sigma}_t}$ (Equation (3)) from the training period serve as the target series for a second NARX network, which learns any nonlinear structure remaining after the GARCH filter and produces out-of-sample predicted residuals $\hat{z}_t$. Third, the final hybrid forecast is reconstructed as

$$\hat{r}_t^{hybrid} = \hat{\mu}_t + \hat{z}_t \cdot \hat{\sigma}_t \quad (7)$$

No further parameter re-estimation is performed at the recombination stage. The implementation proceeds in four stages: (1) data collection, (2) preprocessing, (3) modeling, and (4) evaluation, as summarized in Figure 1.

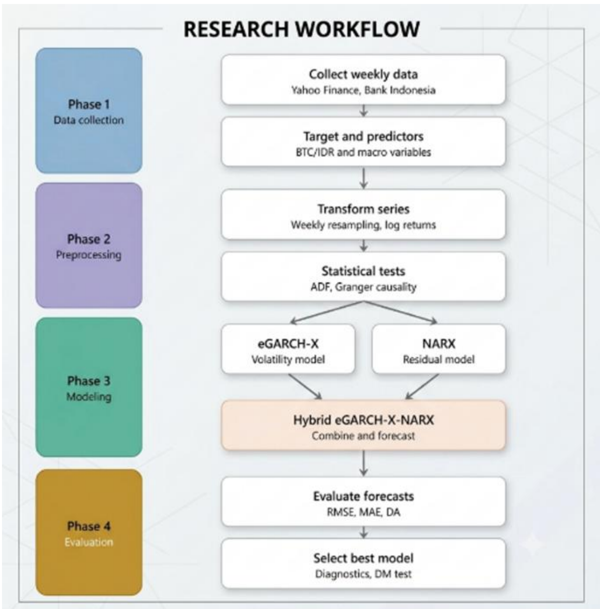

Figure 1. Flowchart of the hybrid eGARCH-X–NARX framework for BTC/IDR forecasting.

2.5 Preliminary Statistical Tests

Prior to estimation, preliminary tests establish the suitability of the data and justify the inclusion of exogenous variables. First, the Augmented Dickey–Fuller (ADF) test in Equation (8) examines the stationarity of each series. The ADF regression is

$$\Delta y_t = \alpha + \beta_t + \gamma y_{t-1} + \sum_{i=1}^{p} \delta_i \Delta y_{t-i} + \varepsilon_t \tag{8}$$

where y_t denotes the series being tested (log returns, and the first difference of the BI rate), Δ is the first-difference operator, t is a deterministic trend, p is the lag order, and ε_t is the error term. The null hypothesis of a unit root (non-stationarity) is tested against the alternative of stationarity. Second, the Granger causality test assesses whether past values of each macroeconomic variable carry predictive information for Bitcoin returns beyond the target's own history, tested at up to 8 lags with the optimal lag per pair selected by minimum F-test p-value, and significance assessed at the 5% level.

2.6 Evaluation Metrics

Out-of-sample forecasting performance is evaluated with three complementary metrics: two error-magnitude measures the Root Mean Square Error (RMSE) in Equation (9), the Mean Absolute Error (MAE) in Equation (10) [1] and a directional measure, the Directional Accuracy (DA) in Equation (11) [15], which is particularly relevant for trading and risk-management decisions, where correctly predicting the sign of price movements matters as much as the error magnitude.

$$RMSE = \sqrt{\frac{1}{n}\sum_{t=1}^{n}(r_t - \hat{r}_t)^2} \tag{9}$$

$$MAE = \frac{1}{n}\sum_{t=1}^{n}|r_t - \hat{r}_t| \tag{10}$$

$$DA = \frac{100}{n-1}\sum_{t=2}^{n} a_t \tag{11}$$

where $a_t = 1$ if $\log r_t \cdot \log \hat{r}_t > 0$ and $a_t = 0$ otherwise.

RMSE penalizes larger errors more heavily, MAE measures the average error magnitude, and DA reports the percentage of correctly predicted upward/downward movements. Lower RMSE and MAE values and a higher DA indicate better forecasting performance. These metrics are used to benchmark the proposed hybrid eGARCH-X-NARX model against the competing forecasting models [1], [22]. Because NARX training depends on random weight initialization, a Diebold Mariano test computed from any single trained network is itself a random draw. Each comparison was therefore repeated across ten independent initializations of the selected architecture. We report the mean and standard deviation of each error metric over those ten runs, the median DM statistic and p-value, and the number of runs attaining significance at the 5% level. Directional accuracy is additionally tested against the chance benchmark of 50% using an exact binomial test on the number of correctly signed forecasts.

3. RESULT AND ANALYSIS

3.1 Descriptive Statistics and Preliminary Tests

The variables consisted of the BTC/IDR log return as the dependent variable, together with the log returns of the IHSG, USD/IDR exchange rate, and gold price, as well as the first difference of the BI Rate as exogenous variables. Log-return transformation was applied to price-based variables, whereas first differencing was used for the BI Rate because an interest rate is not a price variable and, therefore, a log-return transformation is not conceptually appropriate. The four exogenous variables were selected to represent the principal channels through which domestic macro-financial conditions may influence BTC/IDR. The IHSG captures the domestic equity-market and risk-appetite channel; the USD/IDR exchange rate is structurally essential because BTC/IDR is mechanically determined by BTC-USD multiplied by USD/IDR; gold serves as a competing safe-haven asset frequently compared with Bitcoin; and the BI Rate represents the domestic monetary-policy channel. These variables are consistently identified in prior work as determinants of cryptocurrency prices and volatility [15], [17]. Table 2 contains ADF test for BTC/IDR and all exogenous variables.

Figure 2 presents the raw weekly levels of BTC/IDR and the four macroeconomic variables over January 2018–June 2026, with the COVID-19 period shaded. BTC/IDR rose by roughly two orders of magnitude, from about Rp 47 million to a peak near Rp 2.03 billion, exhibiting pronounced non-stationarity that motivates the log-return transformation applied prior to modeling. The IHSG, USD/IDR, and gold series likewise display persistent trends, whereas the BI Rate follows a discrete step pattern that remains constant between policy changes justifying the use of first differencing rather than a log-return transformation for the interest rate. The synchronized disruptions across all series during the COVID-19 period further illustrate that macroeconomic effects on the cryptocurrency market may be concentrated in crisis episodes.

Table 2. ADF test results for BTC/IDR and macroeconomic variables

Variable	ADF Statistics	p-value	lags	Stationary
LR BTC/IDR	-19.9336	0.0000	0	YES
LR IHSG	-20.8570	0.0000	0	YES
LR USD/IDR	-8.9978	0.0000	5	YES
LR Gold USD	-22.4515	0.0000	0	YES
d(BI Rate)	-3.5327	0.0072	12	YES

* LR = Log Return, d = difference

Therefore, the null hypothesis of a unit root was rejected at the 5% significance level for all variables. These results support the use of the transformed series in subsequent time-series modeling and reduce the risk of spurious relationships caused by non-stationary data. The BTC/IDR return distribution was negatively skewed (-0.4908), indicating heavier extreme negative than positive returns, and leptokurtic (excess kurtosis 2.9647). Although not the highest in the sample the first difference of the BI Rate reached 30.9482 because rate changes were infrequent and many weekly values were zero, while the IHSG and USD/IDR showed excess kurtosis of 6.2610 and 5.8491 these values confirm heavier-than-normal tails. The Jarque–Bera test rejected normality for all variables ($p < 0.001$), justifying the use of a skewed Student-t distribution in the eGARCH-X model, since a Gaussian specification would understate the probability of extreme crashes or rallies and thus misestimate volatility and financial risk.

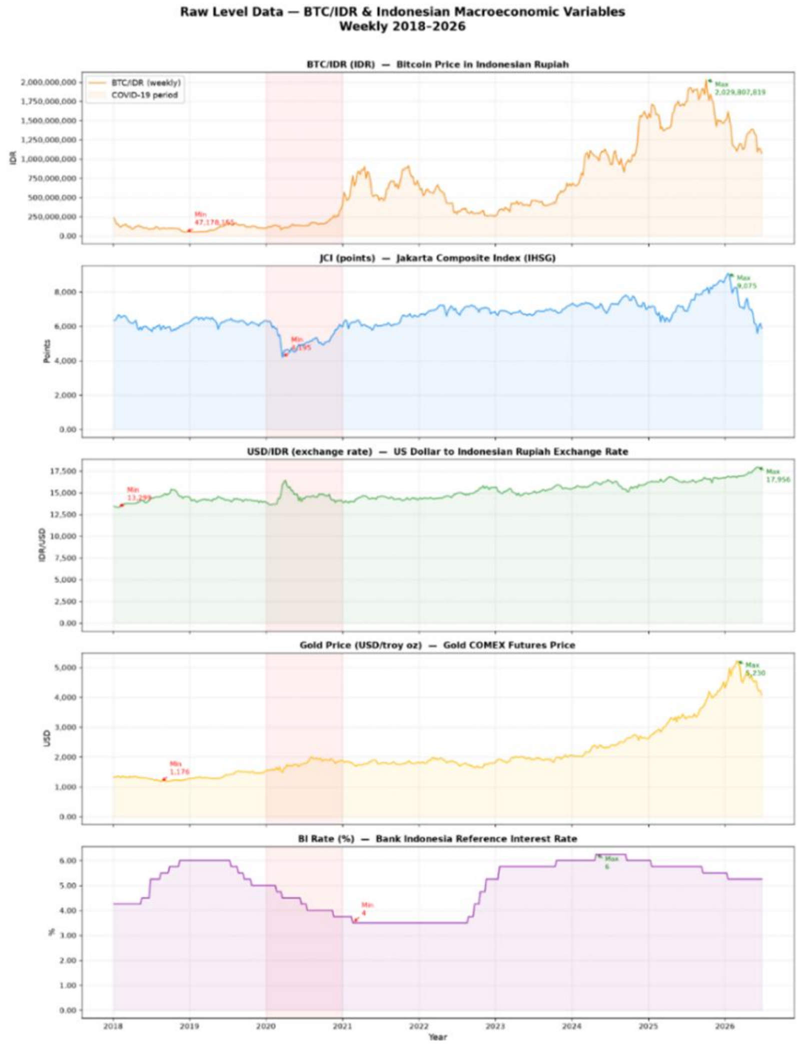

Figure 2. Weekly BTC/IDR price dynamics and macroeconomic variables from January 2018 to June 2026

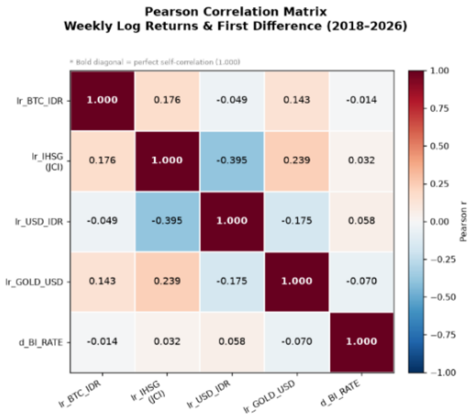

Figure 3. Heatmap correlation for every variables in log return or difference

As shown in Figure 3, the linear relationships between BTC/IDR returns and the domestic macro-financial variables were generally weak, none large enough to indicate a dominant market relationship. The strongest absolute correlation was between the IHSG and USD/IDR (-0.3952), consistent with the tendency of the Indonesian equity market to weaken as the rupiah depreciates. Variance inflation factors confirmed the absence of multicollinearity among the exogenous regressors 1.2349 (IHSG), 1.2008 (USD/IDR), 1.0752 (gold), and 1.0124 (BI Rate) all well below the conventional threshold of 5, indicating that each variable contributed largely independent information to the mean equation.

3.2 Granger Causality Analysis

The Granger causality test was used to determine whether the historical values of each exogenous variable contained additional information for predicting BTC/IDR returns after the past values of BTC/IDR itself had been taken into account. The tests were conducted using a maximum of eight weekly lags. For each test, the null hypothesis stated that the exogenous variable did not Granger-cause the BTC/IDR log return. The results of these tests are reported in Table 3.

The Granger tests and the "-X" terms address different channels: Granger causality is a statement about linear predictability of the conditional mean, whereas the exogenous regressors in Equation (4) enter the conditional variance. Absence of mean predictability therefore does not by itself imply that macro-financial conditions are irrelevant to BTC/IDR volatility. We report both results because they answer different questions, and we do not treat either as a substitute for out-of-sample evidence.

For each Granger test, the lag order was selected using AIC. For USD/IDR, the best specification was obtained at lag six, with an F-statistic of 1.4707 and a p-value of 0.1867. Therefore, the null hypothesis could not be rejected. Gold returns produced an F-statistic of 2.0296 at lag two, with a p-value of 0.1327. The first difference of the BI Rate produced an F-statistic of 1.7270 at lag one and a p-value of 0.1895. Since all four p-values exceeded the 5% significance level, the results did not provide sufficient evidence that USD/IDR, gold, or the BI Rate had incremental predictive power for weekly BTC/IDR returns.

Table 3. Granger causality test results for BTC/IDR returns

Explanatory Variable	Selected lag	F-statistics	p-value	Granger?
LR USD/IDR	6	1.4707	0.1867	NO
LR GOLD USD	2	2.0296	0.1327	NO
d(BI Rate)	1	1.7270	0.1895	NO
LR IHSG	1	1.2132	0.2713	NO

*LR = Log Return, d = difference

These findings are consistent with the weak contemporaneous correlations observed in the descriptive analysis. Granger causality only evaluates predictive information under a particular lag structure and generally within a linear framework. The general absence of significant Granger-causal relationships suggests that weekly BTC/IDR dynamics were more strongly driven by their own historical behavior, changes in volatility regimes, and global cryptocurrency-market factors that were not included in the present model.

Although the Granger tests reveal no linear mean-predictability, a nested likelihood-ratio test comparing eGARCH-X against a restricted eGARCH without exogenous regressors fails to reject the null of zero exogenous effects across all three splits (LR = 7.27, 7.68, and 12.58; df = 8; p = 0.507, 0.465, and 0.127), and both AIC and BIC favor the restricted model. The same conclusion holds out-of-sample: the restricted model attained marginally lower RMSE and MAE and marginally higher directional accuracy at every split, but the Diebold Mariano test found none of these differences significant (p = 0.435, 0.886, 0.835; Table 6). The Granger tests and the "-X" terms nonetheless address different channels Granger causality concerns the conditional mean, whereas the exogenous regressors also enter the conditional variance and we report both because they answer different questions. eGARCH-X is retained as the study benchmark because the hybrid is constructed on its standardized residuals; benchmarking the hybrid against a different first stage would confound the effect of the exogenous terms with the effect of hybridization.

3.3 eGARCH-X Forecasting Results

The eGARCH-X model combined an ARMA conditional mean equation with exogenous regressors and an eGARCH (1,1) variance equation, with ARMA orders selected by AIC-based grid search: ARMA (2,2) for the 80/20 and 85/15 splits (AIC -1.9612 and -1.9928) and ARMA (1,2) for the 90/10 split (AIC -2.1004). The exogenous variables were all insignificant at the 5% level their 95% confidence intervals all contained zero based on Table 4. Results for the 85/15 and 90/10 splits are qualitatively identical.

Table 4. Exogenous coefficient estimates in the conditional mean equation, eGARCH-X (Split 80/20)

Explanatory Variable	Coefficient	95% Confidence Interval
USD/IDR	0.171	[−1.238, 1.581]
Gold	0.442	[−0.186, 1.069]
BI Rate	0.005	[−0.218, 0.227]
IHSG	0.012	[−1.433, 1.457]

The variance equation was considerably stronger: the ARCH parameter (0.072, SE = 0.038, p = 0.011) and the persistence parameter (0.976, SE < 0.001) confirmed pronounced volatility clustering, while the positive asymmetry parameter (0.038, SE = 0.019, p = 0.010) showed that positive shocks raised volatility more than negative ones diverging from the conventional equity-market leverage effect. The estimated shape parameter (4.320, SE = 1.226) and skewness parameter (0.992, SE = 0.074) confirmed heavy-tailed errors, supporting the use of a skewed Student-t over a normal distribution.

The preceding results describe the model's in-sample estimation; out-of-sample forecasting performance is reported separately below. In terms of out-of-sample forecasting performance, eGARCH-X achieved lower RMSE and MAE than both neural specifications at every split (e.g., RMSE 0.0597 and MAE 0.0457 for the 80/20 split, improving to 0.0549 and 0.0414 for the 85/15 split), with residual diagnostics reported in Table 5.

Table 5. Ljung–Box (LB) test for residual of eGARCH-X model

Split	Mean Residual	Std. Residual	LB(12)	LB(5)	Conclusion
80/20	-0.02399	1.0499	7.161<21.03	0.919<11.07	White noise
85/15	-0.0231	1.0334	7.502<21.03	0.912<11.07	White noise
90/10	-0.0128	1.0156	5.344<21.03	3.530<11.07	White noise

Residual diagnostics also supported the adequacy of the eGARCH-X specification. The Ljung–Box statistics for the standardized residuals under the 80/20, 85/15, and 90/10 splits were below the critical value of 21.03. The foregoing estimates identify a conditional variance process that is adequately specified in every respect examined. Volatility exhibits pronounced persistence, responds asymmetrically to the sign of innovations, and is generated by a distribution with substantially heavier tails than the Gaussian benchmark; the standardized residuals satisfy the Ljung–Box criterion at each of the three data splits. The second moment of weekly BTC/IDR returns thus admits a parsimonious and diagnostically adequate representation, and it is in this component, rather than in the conditional mean, that the empirically tractable structure of the series resides.

3.4 NARX Forecasting Results

The NARX model used lagged BTC/IDR returns as feedback inputs and the four exogenous variables as external inputs. The model was trained in open-loop form using teacher forcing and was subsequently converted into a closed-loop architecture to generate recursive forecasts over the testing period. For each split, the configuration achieving the lowest mean validation MSE across its ten runs was retained. The selected specifications were $h = 5, d = 2$ (80/20) and $h = 5, d = 1$ (85/15 and 90/10) for NARX. NARX produced average RMSE of 0.0649, 0.0609, and 0.0680 and average MAE of 0.0469, 0.0429, and 0.0508 (Table 6). Its point-estimate error was higher than eGARCH-X's at every split, but no Diebold Mariano comparison reached significance in any run at the 80/20 and 90/10 splits, and in only one of ten runs at 85/15.

3.5 Hybrid eGARCH-X-NARX Forecasting Results

The hybrid model first used eGARCH-X to estimate the conditional mean and volatility, defined the standardized residual in Equation (3), applied NARX to forecast any remaining nonlinear pattern, and reconstructed the final return forecast in Equation (7). This procedure could improve accuracy only if the eGARCH-X residuals retained systematic nonlinear information; however, the diagnostics showed the residuals were already close to white noise, so NARX received only a weak signal and tended to fit random fluctuations. Direct evidence for this comes from the second stage itself: trained on the standardized residuals, the NARX network produced in-sample predictions of $\hat{z}_t$ that remained essentially flat near zero, indicating that a nonlinear learner explicitly searching for residual structure found none. The hybrid specifications were identical to those of NARX except at the 90/10 split, where $d = 4$ was preferred. The hybrid exceeded eGARCH-X on directional accuracy at the 85/15 and 90/10 splits, but both margins lie within roughly one standard deviation of the run-to-run dispersion and are therefore not meaningful. Performance results for eGARCH-X, NARX and the hybrid summarized in Table 6.

Table 6. Complete evaluation metrics for all models (eGARCH-X, NARX, and Hybrid)

Model	Split	RMSE mean ± std*	MAE ± std*	DA mean ± std* (%)	DM Statistics	p-value	Significant trials
eGARCH	80/20	0.0582	0.0430	47.13	0.7839	0.4350	-
eGARCH-X	80/20	0.0597	0.0457	43.6800	-	-	-
NARX ($h = 5, d = 2$)	80/20	0.0649 ± 0.0043	0.0468 ± 0.0034	55.5200 ± 3.8300	-0.8953	0.3731	0/10
Hybrid ($h = 5, d = 2$)	80/20	0.0644 ± 0.0018	0.0496 ± 0.0013	47.8200 ± 5.7600	-1.6973	0.0946	3/10
eGARCH	85/15	0.0545	0.0397	47.69	0.1430	0.8860	-
eGARCH-X	85/15	0.0549	0.0414	46.15	-	-	-
NARX ($h = 5, d = 1$)	85/15	0.0609 ± 0.0063	0.0429 ± 0.0050	54.6200 ± 2.8300	-0.7044	0.4837	1/10
Hybrid ($h = 5, d = 1$)	85/15	0.0558 ± 0.0025	0.0416 ± 0.0017	49.5400 ± 4.2200	-0.2802	0.5611	0/10
eGARCH	90/10	0.0621	0.0473	46.51	0.2090	0.8350	-
eGARCH-X	90/10	0.0629	0.0493	44.1900	-	-	-
NARX ($h = 5, d = 1$)	90/10	0.0680 ± 0.0045	0.0507 ± 0.0036	44.6500 ± 4.2200	-0.6141	0.5435	0/10
Hybrid ($h = 5, d = 4$)	90/10	0.0795 ± 0.0218	0.0596 ± 0.0119	50.4700 ± 4.6600	-1.2337	0.2249	1/10

*eGARCH-X is the benchmark; DM columns are not defined for its own rows. Neural specifications are reported as mean ± standard deviation over ten independent random initializations; DM and p are medians over the same ten runs, and "sig. trials" counts the runs significant at the 5% level. A negative DM statistic indicates the challenger is less accurate than the benchmark. ARMA orders: (2,2) for 80/20 and 85/15, (1,2) for 90/10; eGARCH(1,1) throughout. No comparison is significant at the 5% level.

3.6 Model Comparison and Diebold Mariano Test

No model separated itself statistically from the others. eGARCH-X produced lower point-estimate RMSE and MAE than both challengers at all three splits, and the restricted eGARCH lower still, but no Diebold Mariano comparison reached significance at the 5% level. Averaged across splits, RMSE was 0.0583 for the restricted eGARCH, 0.0592 for eGARCH-X, 0.0646 for NARX, and 0.0666 for the hybrid an ordering that runs inversely to model complexity, though none of the pairwise differences are significant (Table 6). Directional accuracy was likewise indistinguishable from chance no cell fell outside the non-significant region of a binomial test (all $p > 0.28$), despite a spread of 43.7% to 55.5% across models and splits. Non-significance indicates that the test cannot distinguish the two forecast sequences at conventional levels given the length of the evaluation window; it is not evidence of equivalence, and we therefore do not interpret any challenger as matching the benchmark.

Taken together, the three categories of evidence answer different questions and do not align. In-sample diagnostics (Table 5) confirm that eGARCH-X is adequately specified. Out-of-sample point-forecast accuracy favours the simpler specifications in point estimate, but no Diebold Mariano comparison supports that ordering statistically. Directional accuracy separates no model from chance. The parsimonious specification is therefore preferred on grounds of parsimony rather than demonstrated accuracy: it is not worse, while being deterministic, cheaper to estimate, and directly interpretable.

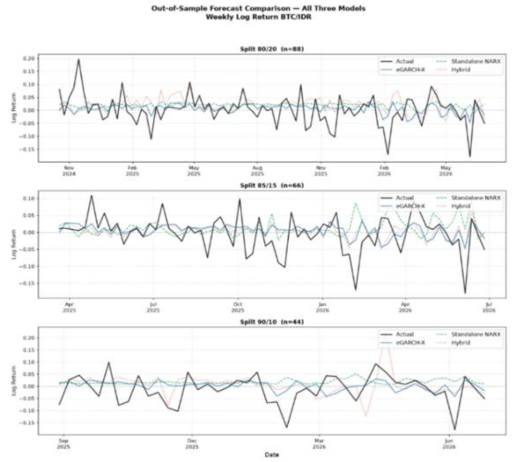

Figure 4. Forecast comparison among eGARCH-X, standalone NARX, and hybrid eGARCH-X–NARX

Figure 4 compares the out-of-sample forecasts of the three models against the actual weekly BTC/IDR log returns across the three data splits. We can conclude that adding a nonlinear component does not automatically improve forecast performance. Hybridization is beneficial only when the first-stage model leaves a systematic residual structure that can be learned. When the standardized residuals already resemble white noise, the additional model may increase forecast variance and the risk of overfitting.

3.7 Implications for the Indonesian Cryptocurrency Market

The results indicate that weekly BTC/IDR dynamics are governed more by evolving volatility regimes than by a stable, predictable conditional mean, consistent with documented cryptocurrency behavior [3]. The estimated persistence (0.9764) implies that shocks affect uncertainty for several subsequent weeks [1], [23] particularly relevant given Bitcoin's markedly higher risk relative to instruments such as the IHSG in the Indonesian market. Investors, exchanges, and risk managers should therefore not treat next week's risk as constant or captured by a long-run historical standard deviation [9]. Directional accuracy, by contrast, ranged from 43.7% to 55.5% across models and splits, and a binomial test rejected chance in no cell (all $p > 0.28$). The practical value of these models therefore lies in the conditional volatility forecast $\hat{\sigma}_t$, which can support position sizing, dynamic stop-loss placement, and value-at-risk limits [1], [15], and whose heavy-tailed innovation distribution matters materially for those applications a Gaussian assumption would systematically understate tail risk. For practitioners in the Indonesian market, these models inform how much exposure to carry, not which direction to take [2].

A central implication concerns model complexity itself. eGARCH-X produced lower point-estimate RMSE and MAE than both challengers at all three splits, and the restricted eGARCH lower still, so that averaged error rose monotonically with model complexity yet no Diebold Mariano comparison supports that ordering statistically. The evidence therefore does not show that the simpler model forecasts better; it shows that the added complexity bought no detectable gain. This finding diverges from much of the recent BTC/USD literature, where hybrid statistical neural frameworks are frequently reported to dominate their individual components [4], [5], [14], underscoring that hybridization gains are conditional rather than general: they materialize only when the base model leaves genuine, learnable structure in its residuals. In this study, the eGARCH-X residuals were already close to white noise, so the second-stage network had no systematic signal to exploit and instead added forecast variance visible in the hybrid's run-to-run RMSE dispersion of 0.0218 at the 90/10 split.

These findings suggest two concrete directions for applied work on cryptocurrency forecasting. First, additive two-stage recombination should be replaced by weighted forecast combination, in which the second-stage weight is estimated rather than fixed at unity, so that a weak residual signal is shrunk toward zero instead of being passed through in full. Second, single-stage estimation, in which the volatility and nonlinear components are estimated jointly, avoids propagating first-stage estimation error into the second-stage target altogether.

Several limitations should be acknowledged and, in turn, define promising directions for future research. First, BTC/IDR was constructed as a synthetic cross-rate (BTC/USD × USD/IDR) raising a potential external-validity concern: exchange-specific liquidity premiums, order-book depth, and local spread dynamics are not directly captured by a globally-quoted cross-rate. To assess the practical severity of this concern, we compared the synthetic series against weekly closing prices from Indodax Indonesia's largest licensed cryptocurrency exchange over a 26-week subsample (January–June 2026). Indodax weekly closing prices for the 26-week validation window were transcribed manually from the exchange's public price chart.

The synthetic construction tracked the exchange-quoted price closely: the mean price-level deviation was near zero (+0.05%, Mean Absolute Deviation (MAD) = 2.05%, std = 2.84%), with absolute deviations exceeding 5% in only 2 of 26 weeks, and the sign of the weekly log return agreed between the two series in 23 of 26 weeks (88.5%). These results indicate no systematic bias in the synthetic construction but it covers only a recent 26-week window rather than the full 2018–2026 sample period. Second, the study excluded global drivers such as sentiment indices, trading volume, funding rates, global equity-market volatility, the US dollar index, and on-chain indicators variables shown elsewhere to carry predictive content for crypto returns and volatility. Third, the neural-network results remained sensitive to architectural choices, training procedures, and recursive forecasting mechanisms, a known characteristic of deep-learning forecasts more broadly [12], [20], [21]. This sensitivity does not overturn the central finding but highlights an additional cost of complexity: greater architectural flexibility also introduces greater instability, compounding the difficulty of achieving reliable gains from hybridization. Fourth, the forecasting protocols are not fully symmetric across model classes. The eGARCH specifications were re-estimated at each step of the rolling out-of-sample window, whereas the NARX networks were trained once and then generated recursive closed-loop forecasts over the entire test period without re-training.

The econometric models therefore incorporate newly realised returns during the evaluation window while the neural models do not. This asymmetry works against the neural specifications and may account for part of their higher point-estimate error; it should be removed in future comparisons by re-training the network on the same rolling schedule. Because the asymmetry favours the econometric specifications, it cannot have produced the null result reported here: even with an informational advantage, eGARCH-X did not attain a significant edge over either challenger.

4. CONCLUSION

This study finds a clear asymmetry between the two components of the weekly BTC/IDR return process. The conditional variance is well identified by a parsimonious eGARCH specification with skewed Student-t innovations, exhibiting high persistence, a significant asymmetry running counter to the conventional equity leverage effect, and heavy tails, with residuals passing all diagnostics across three data splits. The conditional mean is not. Across the three splits, no model demonstrated statistically superior out-of-sample accuracy: Diebold Mariano tests found no significant difference against any challenger, a conclusion stable across ten independent network initializations, and directional accuracy did not differ from chance in any cell. Neither the addition of Indonesian macroeconomic regressors nor a NARX residual stage produced a measurable improvement.

The estimated volatility dynamics remain informative independent of this comparison. Persistence of 0.9764 indicates that shocks affect uncertainty for several subsequent weeks, and the significant positive asymmetry parameter marks a departure from the conventional equity leverage effect. Domestic macroeconomic variables, by contrast, showed limited incremental power for the conditional mean, consistent with the Granger and in-sample likelihood-ratio results reported above.

These results are specific to the present data and specification and do not rule out hybrid gains under a different residual structure. The findings are further qualified by BTC/IDR's construction from BTC/USD and USD/IDR rather than exchange-specific prices, the exclusion of global predictors, and the sensitivity of NARX to architectural choices directions future work may pursue. The parsimonious specification is therefore preferred not because it forecasts the mean better, but because it is not worse while being deterministic, cheaper to estimate, and directly interpretable. Taken together, these results establish a measured upper bound on the mean-predictability of weekly BTC/IDR rather than a new forecasting method: the returns behave close to a martingale, and their modellable content lies in the second moment.

5. REFERENCES

[1] V. Naimy, O. Haddad, G. Fernández-Avilés, and R. El Khoury, "The predictive capacity of GARCH-type models in measuring the volatility of crypto and world currencies," *PLoS One*, vol. 16, no. 1, p. e0245904, 2021, doi: 10.1371/journal.pone.0245904.

[2] Z. Shen, Q. Wan, and D. J. Leatham, "Bitcoin return volatility forecasting: A comparative study between GARCH and RNN," *Journal of Risk and Financial Management*, vol. 14, no. 7, p. 337, 2021, doi: 10.3390/jrfm14070337.

[3] N. A. Kyriazis, "A survey on volatility fluctuations in the decentralized cryptocurrency financial assets," *Journal of Risk and Financial Management*, vol. 14, no. 7, p. 293, 2021, doi: 10.3390/jrfm14070293.

[4] M. Zahid, F. Iqbal, and D. Koutmos, "Forecasting Bitcoin volatility using hybrid GARCH models with machine learning," *Risks*, vol. 10, no. 12, p. 237, 2022, doi: 10.3390/risks10120237.

[5] A. García-Medina and E. Aguayo-Moreno, "LSTM-GARCH hybrid model for the prediction of volatility in cryptocurrency portfolios," *Comput. Econ.*, vol. 63, no. 4, pp. 1511–1542, 2024, doi: 10.1007/s10614-023-10373-8.

[6] Commodity Futures Trading Regulatory Agency (Bappebti), "Cryptocurrency Asset Trading Statistics 2024," Jakarta, Indonesia, 2025.

[7] Q. Phung Duy, O. Nguyen Thi, P. H. Le Thi, H. D. Pham Hoang, K. L. Luong, and K. N. Nguyen Thi, "Estimating and forecasting bitcoin daily prices using ARIMA-GARCH models," *Business Analyst Journal*, vol. 45, no. 1, pp. 11–23, 2024, doi: 10.1108/BAJ-05-2024-0027.

[8] T. Bollerslev, "Generalized autoregressive conditional heteroskedasticity," *J. Econom.*, vol. 31, no. 3, pp. 307–327, 1986, doi: 10.1016/0304-4076(86)90063-1.

[9] D. B. Nelson, "Conditional heteroskedasticity in asset returns: A new approach," *Econometrica*, vol. 59, no. 2, pp. 347–370, 1991, doi: 10.2307/2938260.

[10] H. Han and D. Kristensen, "Asymptotic theory for the QMLE in GARCH-X models with stationary and nonstationary covariates," *Journal of Business & Economic Statistics*, vol. 32, no. 3, pp. 416–429, 2014, doi: 10.1080/07350015.2014.897954.

[11] O. Akgun and E. Gulay, "Dynamics in realized volatility forecasting: Evaluating GARCH models and deep learning algorithms across parameter variations," *Comput. Econ.*, vol. 65, no. 6, pp. 3971–4013, 2024, doi: 10.1007/s10614-024-10694-2.

[12] T. Berger and J. Koubová, "Forecasting Bitcoin returns: Econometric time series analysis versus machine learning," *J. Forecast.*, vol. 43, no. 7, pp. 2904–2916, 2024, doi: 10.1002/for.3158.

[13] F. U. Khan, F. Khan, and P. A. Shaikh, "Forecasting returns volatility of cryptocurrency by applying various deep learning algorithms," *Future Business Journal*, vol. 9, no. 1, p. 25, 2023, doi: 10.1186/s43093-023-00200-9.

[14] W. Khoso and M. R. Doudpoto, "Forecasting bitcoin price volatility using GARCH models and real-time data," *Int. J. Cryp. Curr. Res.*, vol. 5, no. 2, pp. 28–41, 2025, doi: 10.51483/IJCCR.5.2.2025.27-41.

[15] S. A. Basher and P. Sadorsky, "Forecasting Bitcoin price direction with random forests: How important are interest rates, inflation, and market volatility?," *Machine Learning with Applications*, vol. 9, p. 100355, 2022, doi: 10.1016/j.mlwa.2022.100355.

[16] S. A. H. Havidz, V. E. Karman, and I. Y. Mambea, "Is Bitcoin price driven by macro-financial factors and liquidity? A global consumer survey empirical study," *Organizations and Markets in Emerging Economies*, vol. 12, no. 2, pp. 399–414, 2021, doi: 10.15388/omee.2021.12.62.

[17] S. Peng, C. Prentice, S. Shams, and T. Sarker, "A systematic literature review on the determinants of cryptocurrency pricing," *China Accounting and Finance Review*, vol. 26, no. 1, pp. 1–30, 2024, doi: 10.1108/CAFR-05-2023-0053.

[18] E. Bouri, P. Molnár, G. Azzi, D. Roubaud, and L. I. Hagfors, "On the hedge and safe haven properties of Bitcoin: Is it really more than a diversifier?," *Financ. Res. Lett.*, vol. 20, pp. 192–198, 2017, doi: 10.1016/j.frl.2016.09.025.

[19] J. Klose, "Comparing cryptocurrencies and gold: a system-GARCH-approach," *Eurasian Economic Review*, vol. 12, no. 4, pp. 653–679, 2022, doi: 10.1007/s40822-022-00218-4.

[20] T. Lin, B. G. Horne, P. Tino, and C. L. Giles, "Learning long-term dependencies in NARX recurrent neural networks," *IEEE Trans. Neural Netw.*, vol. 7, no. 6, pp. 1329–1338, 1996, doi: 10.1109/72.548162.

[21] A. H. Dhafer *et al.*, "Empirical analysis for stock price prediction using NARX model with exogenous technical indicators," *Comput. Intell. Neurosci.*, vol. 2022, p. 9208640, 2022, doi: 10.1155/2022/9208640.

[22] V. Y. Naimy and M. R. Hayek, "Modelling and predicting the Bitcoin volatility using GARCH models," *International Journal of Mathematical Modelling and Numerical Optimisation*, vol. 8, no. 3, pp. 197–215, 2018, doi: 10.1504/IJMMNO.2018.088994.