

Do Macroeconomic Variables Improve Rupiah Forecasts under Conditional Heteroskedasticity?

Sry Yulianti Lobo ¹, Rahmat Syam ^{2*}, Wahidah Sanusi ³

^{1,2,3} Master's Program in Mathematics, Faculty of Mathematics and Natural Sciences, Universitas Negeri Makassar

Jl. Daeng Tata Raya, Parangtambung, Makassar 90224, Indonesia

250027301004@student.unm.ac.id ¹, rahmat.syam@unm.ac.id ², wahidah.sanusi@unm.ac.id ³

* Corresponding author

Article Info

Article history:

Received 2026-09-16

Revised 2026-10-06

Accepted 2026-10-19

Keyword:

ARIMAX-GARCH,
Exchange Rate Forecasting,
Interval Forecasting,
Meese-Rogoff Puzzle,
Rolling-Window Evaluation.

ABSTRACT

Rupiah exchange-rate forecasting studies are usually univariate and rarely evaluated under a leakage-free protocol against a strong naive benchmark. This study tests whether three macroeconomic variables (Federal Funds Rate, Bank Indonesia policy rate, and Indonesian Consumer Price Index) improve forecasts from a linear ARIMA-GARCH model, whose GARCH component captures conditional heteroskedasticity. The resulting ARIMAX-GARCH is compared with a nested ARIMA-GARCH control and random walks with and without drift. Using daily JISDOR data (2018 to 2026), models are estimated by quasi-maximum likelihood in 1,000-day rolling windows at horizons of one day to one quarter. Point forecasts are evaluated by MAE, RMSE, and MAPE with Clark-West and Giacomini-White tests, and interval forecasts by coverage, the interval score, and the Kupiec and Christoffersen tests. The macroeconomic variables add no predictive value over the nested control at any horizon under either test. This null result persists under an error-correction extension, Student-t errors, alternative window sizes, EGARCH and GJR-GARCH variances, and two separate subperiods, consistent with the Meese-Rogoff puzzle within the daily linear specification tested. The model beats the random walk without drift, but only because of the depreciation trend, since the random walk with drift reproduces about 97% of the gain. The GARCH component adds value in interval rather than point forecasting, giving well-calibrated 95% intervals about 12% narrower than constant-variance benchmarks at one day. Macroeconomic variables therefore do not improve daily rupiah forecasts from ARIMA-GARCH models, and gains over a random walk should be tested against a drift benchmark.



This is an open access article under the [CC-BY-SA](https://creativecommons.org/licenses/by-sa/4.0/) license.

I. INTRODUCTION

The rupiah exchange rate against the United States dollar is one of the most closely watched macro-financial variables in the Indonesian economy. Its movements shape the monetary policy and foreign-exchange interventions of Bank Indonesia, the budget assumptions of the Ministry of Finance, the hedging decisions of firms exposed to foreign currency, and the asset allocation of investors. The need for accurate forecasts has grown as the rupiah weakened to a record low of about IDR 18,171 per US dollar in June 2026 [1]. The rupiah series has three statistical properties that determine how it should be modeled. Its level is nonstationary and carries a unit root, which appears as a persistent upward trend.

Its returns are conditionally heteroskedastic, with periods of high and low volatility that tend to cluster. It also responds to exogenous shocks transmitted through monetary policy and inflation. These properties motivate a framework that differences the series, models the conditional variance explicitly, and admits exogenous macroeconomic inputs, which is the ARIMAX-GARCH specification.

A large body of work has forecast the rupiah exchange rate, from classical statistical approaches such as ARIMA and GARCH to machine learning and deep learning models such as the extreme learning machine, gradient boosting, and long short-term memory, which commonly report a mean absolute percentage error below 2% [2][3][4][5][6][7][8]. A critical

review, however, reveals two gaps that remain unaddressed. The first gap is informational. Most rupiah studies rely on the historical exchange-rate series as the only input, without the macroeconomic variables that theory [9] and empirical work [10][11] recognize as determinants of exchange-rate dynamics, and none has combined such variables with explicit conditional-variance modeling as in the ARIMAX-GARCH specification. The studies that do relate the rupiah to macroeconomic variables [10][11] estimate in-sample explanatory models rather than evaluating out-of-sample forecasts. A transparent ARIMA-GARCH model succeeds on the weekly rupiah with a test MAPE of 1.99% [12], but specifications with exogenous predictors appear only outside the rupiah context, for the Chinese yuan [13], the Indian rupee [14], the Turkish lira [15], the Egyptian pound [16], and the United States dollar with GARCH-X errors estimated by quasi-maximum likelihood [17]. Their evidence is mixed. Economic features improve deep learning forecasts of the yuan [13], whereas exogenous variables help the official Egyptian rate but not the parallel-market rate [16]. Whether macroeconomic variables improve rupiah forecasts within a model that accounts for conditional heteroskedasticity has therefore not been examined.

The second gap concerns the evaluation method. The exchange rate operates similarly to a random walk, which has proven challenging to surpass since the work of Meese and Rogoff [18]; thus, a minimal one-step MAPE generated by a model that simply reiterates the prior observation does not inherently demonstrate predictive capability. Rupiah studies generally use a static train-test split rather than a rolling window that preserves temporal ordering, and even those comparing ARIMA, LSTM, and ensemble models report a very small MAPE without a random-walk benchmark [3][19]. When a naive benchmark is used, it is almost always a random walk without drift. For a strongly trending series such as the rupiah, a random walk with drift is the more demanding benchmark, and without it an advantage that stems entirely from the trend can be mistaken for forecasting ability. Stricter protocols exist outside the rupiah context, including the honest reporting of a negative coefficient of determination [15], validation that preserves temporal ordering [20], rolling-window volatility evaluation [21], and rolling-origin evaluation of Indonesian inflation forecasts, which finds that complex models do not significantly beat naive forecasts [22]. Rupiah studies, moreover, focus on point accuracy, although the main purpose of modeling the conditional variance is to quantify uncertainty, so the value of a GARCH component is best judged through interval forecasts [23][24].

The primary objective of this study is therefore to test whether three macroeconomic variables, the Federal Funds Rate (FFR), the Bank Indonesia policy rate (BI rate), and the Indonesian Consumer Price Index (CPI), which represent the interest-rate and inflation channels of exchange-rate theory [9] (Section II.A), improve rupiah forecasts from a linear ARIMA-GARCH model. The resulting ARIMAX-GARCH model is compared with a nested ARIMA-GARCH control

that shares the same autoregressive and GARCH dynamics but omits the macroeconomic regressors, so any difference in accuracy is attributable to the macroeconomic variables alone. Four supporting analyses make the answer credible. First, both models are compared with a random walk without drift and a random walk with drift, which reveals whether any advantage reflects genuine predictability or only the depreciation trend. Second, an error-correction extension (ARIMAX-ECM-GARCH) tests whether the macroeconomic variables contribute through a long-run equilibrium channel that differencing would remove. Third, 95% interval forecasts are evaluated to test whether the conditional-variance component adds value beyond point accuracy. Fourth, the main comparison is repeated under alternative error distributions, window lengths, and asymmetric variance specifications. All forecasts are produced under a leakage-free, fixed-size rolling-window protocol, and differences in point accuracy are assessed with two complementary tests for nested models. The scope is deliberately restricted to the linear ARIMA-GARCH family, so that the contribution of the macroeconomic variables is isolated within otherwise identical models. Comparison with machine learning and deep learning models is left for future work.

This study makes two contributions. The first is methodological. It supplies a reproducible, leakage-free rolling-window protocol with tiered naive benchmarks and a nested control, which together separate genuine predictability from the depreciation trend. The second is empirical. This is, to the best of the authors' knowledge, the first study to test whether macroeconomic variables improve rupiah forecasts within a model that explicitly accounts for conditional heteroskedasticity, evaluated on both point and interval accuracy. The ARIMAX-GARCH specification itself is standard and is not claimed as a contribution. The model is therefore the object of study, while the protocol is the instrument that makes the test fair.

II. METHOD

A. Data and Variables

The data consist of the daily rupiah exchange rate and three exogenous macroeconomic variables. The forecast target is the Jakarta Interbank Spot Dollar Rate (JISDOR) published by Bank Indonesia [1], covering the period from 2 April 2018 to 27 July 2026. The JISDOR is published only on Indonesian trading days, so the series is in trading-day time and contains no calendar gaps to impute. Non-trading days such as weekends and public holidays have no fixing and are absent from the sample rather than interpolated. The aligned sample contains 2,004 daily observations with no missing values, which yield 2,003 daily log returns after differencing, and 27 July 2026 is the data cutoff used throughout. The series is indexed in trading-day time, so changes in the trading calendar across the sample, such as added or moved public holidays, require no adjustment. The exogenous regressors are the FFR, measured as the upper limit of the Federal Reserve's target range (FRED series DFEDTARU) [25], the

BI rate [26], and the Indonesian CPI, reconstructed by chaining the monthly month-to-month inflation rates published by Statistics Indonesia (BPS) [27]. The BI rate is the main policy instrument through which Bank Indonesia conducts its monetary policy mix [28], while the FFR and the CPI capture external monetary conditions and domestic price pressure. Together they form the exogenous vector $W_t = (i_t^{\text{FFR}}, i_t^{\text{BI}}, \text{CPI}_t)'$, where i_t^{FFR} is the Federal Funds Rate, i_t^{BI} is the BI rate, and CPI_t is the Consumer Price Index on trading day t . The three variables are treated as predetermined with respect to the exchange rate X_t , which means that their values on trading day t are fixed before X_t is realized, so that $E(\varepsilon_t | W_t, W_{t-1}, \dots) = 0$.

The selection of the Federal Funds Rate, the Bank Indonesia policy rate, and the Indonesian Consumer Price Index as exogenous variables rests on established exchange rate theory rather than on data mining. The Federal Funds Rate and the Bank Indonesia policy rate together form the interest differential that uncovered interest parity places at the center of exchange rate determination, because under this condition the expected depreciation of the rupiah equals the gap between the domestic and foreign nominal interest rates [9]. The Indonesian Consumer Price Index captures the inflation channel identified by relative purchasing power parity, since a higher domestic price level relative to the United States should depreciate the rupiah over the medium run [9]. Exchange rate pass-through supports treating these variables jointly, because movements in the rupiah feed back into domestic prices and make the link between the exchange rate and its fundamentals bidirectional [28]. The Indonesian empirical record justifies including these fundamentals while cautioning against strong expectations of predictive gain, since Tang finds that the precise form of uncovered interest parity is strongly rejected for Indonesia [29], and Fianto and coauthors fail to find evidence that macroeconomic fundamentals predict rupiah returns over a long historical sample, while inflation and the US Treasury bill rate become significant predictors only during the Asian financial crisis [30]. This combination of theoretical grounding and mixed empirical support motivates the present design, which treats the three variables as economically justified regressors whose practical forecasting contribution must be established out of sample rather than assumed.

B. Preprocessing and Frequency Alignment

The three exogenous variables are observed at a lower frequency than the daily exchange rate. The CPI is released monthly with a publication lag, while the BI rate and the FFR are set at scheduled meetings and remain unchanged between meetings. The exogenous series are aligned to the daily frequency of X_t by last-observation-carried-forward, so that each policy rate is held constant on every trading day until the next announcement, and each CPI figure is mapped to all trading days after its release date. This step-function alignment on a daily grid is exact for the two policy rates,

which are legally constant between announcement dates, and adequate for the CPI, which has a single value per month.

The spot rupiah series is empirically integrated of order one, $I(1)$, so the working series for the conditional mean is the first difference of the logarithm, that is the log return $r_t = 100 \Delta \ln X_t$ in percent, which is stationary, $I(0)$. The ARIMAX-GARCH specification operates directly on this return series, together with the first differences of the exogenous regressors in their original units. The relevant source of leakage in a time-series context is temporal, that is when information after the forecast origin is allowed to influence estimation. This is prevented by the fixed-size rolling-window protocol described in Section II.E, under which all coefficients, including the cointegrating vector of the error-correction extension, are re-estimated using only information up to the forecast origin. The unit-root and cointegration pre-tests in Section II.C use the full sample, but they determine only the transformation of the series, not any coefficient used for forecasting. At every horizon, including $h = 1$, the future differences of the exogenous regressors are set to zero, so the macroeconomic variables affect the forecast only through the coefficients estimated within each window.

C. Stationarity and Cointegration Testing

The order of integration of each series is determined using the Augmented Dickey-Fuller (ADF) and Phillips-Perron (PP) tests applied in parallel, with MacKinnon critical values [31]. Both tests share the same null hypothesis, namely the presence of a unit root, and both include a constant term. The ADF test augments the regression with lagged differences whose number is selected by the AIC, while the PP test applies a nonparametric Newey-West correction that is consistent under heteroskedasticity and autocorrelation. Using both tests together provides cross-confirmation that is well suited to the conditionally heteroskedastic nature of the exchange-rate series.

Whether the exchange rate and the regressors share a common stochastic trend is examined with the Johansen cointegration test, using the trace and maximum-eigenvalue statistics [32], applied to the log exchange rate and the three exogenous variables in levels. The test includes a constant term, and the lag order of the underlying vector autoregression is selected by the AIC from a maximum of 10 lags. Both the unit-root and the cointegration tests are applied to the full sample, as explained in Section II.B.

D. The ARIMAX-GARCH Model

The ARIMAX model extends ARIMA by including the exogenous vector W_t as a predictor of the conditional mean, as shown in equation (1),

$$\varphi(L) \Delta^d \ln X_t = c + \xi' W_t + \theta(L) \varepsilon_t, \quad (1)$$

where L is the lag operator, $\varphi(L)$ and $\theta(L)$ are the autoregressive and moving-average polynomials, $\Delta = (1 - L)$ is the differencing operator with integration order d , and $\xi = (\xi_1, \dots, \xi_k)'$ is the vector of regression coefficients. The

order of integration on the two sides of the equation must be balanced. Each I(1) exogenous variable is therefore differenced, because the left-hand side has been made I(0) by Δ^d , so that spurious regression is avoided and equation (1) becomes (2).

$$\varphi(L) \Delta^d \ln X_t = c + \xi' \Delta W_t + \theta(L) \varepsilon_t \quad (2)$$

When the identification indicates a moving-average component, that component is approximated by a higher-order autoregressive representation. With $d = 1$ and the AR(2) order selected as described in Section II.E, the conditional mean estimated in this study is

$$r_t = c + \varphi_1 r_{t-1} + \varphi_2 r_{t-2} + \xi' \Delta W_t + \varepsilon_t, r_t = 100 \Delta \ln X_t \quad (3)$$

where $\Delta W_t = (\Delta i_t^{\text{FFR}}, \Delta i_t^{\text{BI}}, \Delta \text{CPI}_t)'$ and ε_t is the mean-equation error whose conditional variance is modeled below.

Volatility clustering means that the conditional variance σ_t^2 is not constant but a measurable function of the information set $\mathcal{F}_{t-1}$, a feature first formalized in the ARCH model of Engle [33] and generalized to the GARCH family by Bollerslev [34], whose predictive capacity for currency volatility has been confirmed across both cryptocurrencies and world currencies [35]. The mean-equation error is decomposed as $\varepsilon_t = \sigma_t z_t$, where z_t is independent and identically distributed with zero mean and unit variance. Its conditional variance follows a GARCH(p, q) process as in equation (4),

$$\sigma_t^2 = \alpha_0 + \sum_{i=1}^q \alpha_i \varepsilon_{t-i}^2 + \sum_{j=1}^p \beta_j \sigma_{t-j}^2, \quad (4)$$

with $\alpha_0 > 0$, $\alpha_i \geq 0$, $\beta_j \geq 0$. The process is covariance-stationary if and only if $\sum \alpha_i + \sum \beta_j < 1$, which for GARCH(1,1) reduces to $\alpha_1 + \beta_1 < 1$. The ARIMAX conditional-mean component (3) and the GARCH conditional-variance component (4) are combined into a single ARIMAX-GARCH system and estimated simultaneously. Joint estimation is required because the mean-equation error ε_t has a conditional variance σ_t^2 that depends on $\mathcal{F}_{t-1}$, so that separate estimation would yield inefficient estimates of the mean parameters. Estimation maximizes the conditional log-likelihood, which under conditional normality and up to a constant is given by equation (5).

$$\ell_T(\theta) = -\frac{1}{2} \sum_{t=1}^T \left(\ln \sigma_t^2 + \frac{\varepsilon_t^2}{\sigma_t^2} \right) \quad (5)$$

Empirical exchange-rate data generally violate the normality assumption, so estimation uses Quasi-Maximum Likelihood Estimation (QMLE). This approach was recently applied to a GARCH-X specification for exchange-rate forecasting by Fakhri and AlMohana [17]. Maximizing the Gaussian likelihood still yields consistent estimates provided the mean and variance equations are correctly specified. When the series are cointegrated, as tested in Section II.C, full differencing in equation (1) removes the long-run equilibrium information contained in the levels.

This study extends the ARIMAX-GARCH model to a single-equation error-correction form, the ARIMAX-ECM-

GARCH, to test whether that equilibrium channel provides additional predictive value, as in equation (6).

$$\Delta \ln X_t = c + \varphi(L) \Delta \ln X_{t-1} + \xi' \Delta W_t + \lambda z_{t-1} + \varepsilon_t, \quad \varepsilon_t | \mathcal{F}_{t-1} \sim \text{GARCH}(1,1) \quad (6)$$

The term $z_{t-1} = \ln X_{t-1} - (a + \beta' W_{t-1})$ is the equilibrium error, that is the deviation of the exchange rate from its long-run relationship with the FFR, the BI rate, and the CPI, while λ is the speed of adjustment, expected to be negative. The Engle-Granger long-run regression [36] estimates the cointegrating vector (a, β) in levels within each rolling window. The FFR, the BI rate, and the CPI are treated as weakly exogenous with respect to the cointegrating vector [37]. This property makes the single-equation specification adequate while preserving the univariate GARCH component. The model is nested within the ARIMAX-GARCH model and becomes identical to it when $\lambda = 0$. A Clark-West test of this model therefore examines directly whether the long-run equilibrium channel genuinely adds point-forecast accuracy.

E. Model Construction and Rolling-Window Protocol

The rolling-window protocol is defined as follows [21], [38]. Each estimation window contains 1,000 consecutive trading-day log returns, producing 1,003 successive forecast origins over the 2018 to 2026 sample (944 at the quarterly horizon). The first window spans 3 April 2018 to 19 May 2022 and serves only for estimation, so the out-of-sample period runs from 20 May 2022, the first one-day target, to 27 July 2026. At each origin, the previously identified orders are held fixed and only the coefficients are re-estimated on the in-window data, so re-estimation occurs at every one-step shift of the window. Forecasts are generated for $h = 1, 5, 10, 20$, and 60 trading days (one day to one quarter) by iterating the estimated mean recursion forward. The window then advances by one observation and the procedure repeats. All benchmarks (random walk without drift, random walk with drift, and the nested ARIMA-GARCH control) use the identical windows, origins, and information set, so any accuracy difference arises only from specification and not from a different estimation period, window length, or information available at the forecast origin. Future values never enter estimation, because each window ends at its origin and every exogenous regressor is held at its last observed value when forecasts are iterated beyond the origin. The window size is fixed rather than expanding, in view of possible structural change in the rupiah dynamics, which makes distant observations less representative of the current data-generating process.

The model is built through the following steps. First, the orders of integration and differencing are set according to Section II.C and equation (1). Second, candidate autoregressive orders are identified from the sample ACF and PACF of the log returns and then compared, for autoregressive orders from one to three and GARCH orders (1,1), (1,2), and (2,1), by the AIC and BIC, subject to residual

adequacy measured by the Ljung-Box test on the standardized residuals. This is carried out in two windows: the first estimation window, which ends before any forecast origin, and the most recent window. The selected orders are then frozen and not re-identified at each rolling-window iteration, so that only their coefficients are re-estimated per window. The same orders are imposed on the ARIMAX-GARCH model and the nested control, so the order choice cannot favor the macroeconomic variables. The ARIMAX-ECM model re-estimates the cointegrating vector ($\alpha \beta$) on the in-window data at each iteration, so the equilibrium error z_{t-1} is always formed from information up to the forecast origin. Third, the mean-equation residuals are examined with the ARCH-LM test to confirm the conditional heteroskedasticity that motivates the GARCH component. Fourth, the mean and variance parameters are estimated jointly by a single QMLE. Fifth, the covariance-stationarity condition of the GARCH process is verified. Sixth, the standardized residuals are checked with the Ljung-Box test for remaining autocorrelation at lag 20, about one trading month, the ARCH-LM test for remaining ARCH effects at lag 10, about two trading weeks, and the Jarque-Bera test for normality.

F. Benchmark Models

The random walk (RW) without drift, which forecasts no change, is the naive lower bound, since it tends to match or outperform structural exchange-rate models at short horizons under weak-form market efficiency. The RW ignores, however, the persistent depreciation of the rupiah, so a model with a nonzero mean return can beat it simply by extrapolating that trend. The RW with drift is therefore included as a stricter benchmark, with the h -step level forecast $\hat{X}_{o+h} = X_o e^{h\hat{\mu}_o}$, where $\hat{\mu}_o$ is the mean log return of the current window. The nested ARIMA-GARCH control has the same AR(2) mean and GARCH(1,1) variance as the full specification but omits W_t , so it is the ARIMAX-GARCH model with the regression coefficients set to zero, and its comparison with the full model isolates the contribution of the macroeconomic regressors. The ARIMAX-ECM-GARCH is evaluated against the same benchmarks. All benchmarks follow the protocol of Section II.E, and all forecasts are converted to the level scale for the accuracy metrics, while significance is tested on the log-return scale (Section II.G).

G. Evaluation Metrics and Significance Testing

Forecasts are evaluated on two dimensions, point accuracy and interval quality, both under the same leakage-free rolling-window protocol. Forecast accuracy at each iteration is measured from the error $e_i = y_i - \hat{y}_i$ using three standard point-forecast metrics [38], the Mean Absolute Error (MAE), the Root Mean Squared Error (RMSE), and the Mean Absolute Percentage Error (MAPE), defined as $MAE = \frac{1}{n} \sum_{i=1}^n |e_i|$, $RMSE = \sqrt{\frac{1}{n} \sum_{i=1}^n e_i^2}$, and $MAPE = \frac{100\%}{n} \sum_{i=1}^n \left| \frac{e_i}{y_i} \right|$. The three metrics are used together because each highlights a different aspect of the error. The MAE measures the typical error magnitude and is robust to outliers, the RMSE captures sensitivity to sharp spikes caused by macroeconomic shocks to the rupiah, and the MAPE provides a scale-free proportional measure.

A smaller MAE, RMSE, or MAPE does not by itself guarantee a significant advantage, since such a difference can arise from sampling variability alone. A significance test is therefore required. All comparisons in this study are nested. The RW without drift is a special case of the RW with drift, which is in turn nested within the ARIMA-GARCH, ARIMAX-GARCH, and ARIMAX-ECM-GARCH models through the restriction of some coefficients to zero. This study uses the Clark-West statistic, which is designed specifically for nested comparisons [39].

The Clark-West statistic corrects the excess MSPE of the larger model that arises from estimating parameters equal to zero. At each forecast origin it is defined as in equation (7),

$$f_t = e_{small,t}^2 - [e_{large,t}^2 - (\hat{y}_{small,t} - \hat{y}_{large,t})^2] \quad (7)$$

where $e_{small,t}$ and $e_{large,t}$ are the errors of the parsimonious model and its nesting model. The test statistic is the t -ratio of the mean of f_t against a HAC Newey-West standard error [40] with lag order $h-1$. The hypothesis is one-sided, namely $H_0: E[f_t] \leq 0$ against $H_1: E[f_t] > 0$, with a standard-normal critical value of 1.645 at the 5% level. The rolling-window scheme yields a test that is better sized, so this normal approximation is appropriate for the present protocol [39]. Power nonetheless declines at long horizons, because the multi-step forecasts overlap and the number of origins shrinks to 944 at the quarterly horizon, so the p -values at $h = 20$ to $h = 60$ should be read with greater caution. As a complement that targets finite-sample predictive ability rather than population-level parameter relevance, this study also reports the unconditional Giacomini-White test of equal predictive ability, since it remains valid for nested models under the fixed-size rolling window used here [41]. The test is applied to the loss differential $d_t = e_{small,t}^2 - e_{large,t}^2$ of the squared log-return errors, where a positive mean favors the larger model, and its variance is estimated by the Newey-West HAC estimator with lag order $n^{1/3}$ rounded to the nearest integer, which equals 10 for the number of forecast origins in this study. Since each comparison is tested at five horizons, the p -values are adjusted using the Holm procedure [42], which controls the family-wise error rate across horizons.

Forecasts are produced in log-return space and converted to the level before evaluation. For a horizon h , the level forecast is given by equation (8),

$$\hat{X}_{t+h} = X_t \exp\left(\frac{1}{100} \sum_{k=1}^h \hat{r}_{t+k}\right) \quad (8)$$

where $\hat{r}_{t+k}$ are the iterated log-return forecasts in percent. Taking logarithms shows that the log-level forecast error is the sum of the h return-forecast errors, $\ln X_{t+h} - \ln \hat{X}_{t+h} = \frac{1}{100} \sum_{k=1}^h (r_{t+k} - \hat{r}_{t+k})$, so the errors made at each step accumulate rather than cancel, and the variance of the level error grows approximately linearly with h when the return errors are weakly correlated. This explains the monotone growth of the level MAE and RMSE with h . The exponential transformation also introduces a retransformation bias of the Jensen type, of order one half the forecast variance, which can create a spurious advantage of a fraction of a percent [38]. Significance is therefore tested in log-return space, where the log-return error is $e_t = 100 [\ln(y_t) - \ln(\hat{y}_t)]$ and the nested structure remains exact and linear.

Beyond point accuracy, the quality of interval forecasts is also assessed in order to test the usefulness of the conditional-variance component. At each origin and each horizon, a 95% prediction interval is formed under the same rolling-window protocol as the point evaluation, so there is no information leakage. For the GARCH-based models, namely the ARIMA-GARCH, ARIMAX-GARCH, and ARIMAX-ECM-GARCH, the h -step forecast distribution is not available in closed form because of the autoregressive dynamics and the changing variance, so the interval is formed by path simulation. At each iteration, 2,000 forward return paths are simulated, with the conditional variance updated at each step through the GARCH recursion in equation (4), using the parameters estimated in the current window. Each path is converted to levels by cumulative exponentiation, and the interval bounds are taken from the 2.5th and 97.5th percentiles of the simulated level distribution. The simulated innovations are drawn from the standard normal distribution, consistent with the QMLE, and the adequacy of this choice is assessed through the empirical coverage. The differences of the future exogenous variables are set to zero, so no future macroeconomic value is read.

Three constant-variance benchmarks are included in order to isolate the source of interval quality. The RW and the RW with drift use the sample variance $\hat{\sigma}^2$ of the log returns in the current window, so their 95% interval, $X_{\text{exp}}(h\hat{\mu}_o \pm 1.96\sqrt{h\hat{\sigma}^2})$, widens with the square root of the horizon but does not react to recent volatility, where $\hat{\mu}_o = 0$ for the RW without drift. As an ablation control, an ARIMAX model with its variance frozen at the unconditional GARCH variance, $\hat{\sigma}^2 = \omega/(1 - \alpha_1 - \beta_1)$, computed from the ARIMAX-GARCH estimates of the current window, is also included. It retains the full mean equation but strips out the changing-variance property, so its difference from the ARIMAX-GARCH model measures the added value of the conditional variance directly.

Interval quality is assessed through three aspects. Empirical coverage compares the proportion of realizations that fall inside the interval against the nominal level of 95%. The mean interval width measures sharpness, and at equal coverage a narrower interval is better because it indicates smaller uncertainty. The interval score combines the two by penalizing both width and violations [23]. Calibration is tested formally through the Kupiec test for unconditional coverage [43] and the Christoffersen test for conditional coverage, which also checks the independence of violations [24]. Because the multi-step forecasts overlap, the Christoffersen independence test is interpreted only at the one-day horizon, while at longer horizons the assessment rests on coverage, width, and the interval score.

H. Robustness Checks

The comparison between the ARIMAX-GARCH model and the nested ARIMA-GARCH control is repeated under five alternative specifications, giving six specifications in total including the baseline, so that the conclusion on the macroeconomic contribution can be shown not to depend on particular modeling choices. Each alternative changes one element while the rest of the protocol is kept fixed, and both models in each pair always share the same specification. The first replaces the Gaussian quasi-likelihood with a Student-t likelihood whose degrees of freedom are estimated within

each window, which accommodates the heavy tails found in the residual diagnostics [44]. The second and third change the rolling-window length from 1,000 to 750 and 1,250 observations. These two specifications begin forecasting after their own first full window, so they are evaluated over longer and shorter out-of-sample periods than the baseline, and each is compared only with its own nested control. The fourth and fifth replace the GARCH(1,1) variance with asymmetric specifications that allow positive and negative shocks to affect volatility differently. The EGARCH(1,1) model [45] specifies the logarithm of the conditional variance as shown by equation (9),

$$\ln \sigma_t^2 = \omega + \alpha(|z_{t-1}| - E|z_{t-1}|) + \gamma z_{t-1} + \beta \ln \sigma_{t-1}^2 \quad (9)$$

so no positivity constraint on the parameters is required, while the GJR-GARCH(1,1) model [46] as in equation (10) adds a leverage term,

$$\sigma_t^2 = \omega + (\alpha + \gamma I_{t-1})\varepsilon_{t-1}^2 + \beta \sigma_{t-1}^2 \quad (10)$$

where $I_{t-1} = 1$ if $\varepsilon_{t-1} < 0$ and zero otherwise. The macroeconomic contribution in each specification is measured as the percentage reduction in level RMSE of the ARIMAX-GARCH model relative to its nested control, and it is tested with the Clark-West and Giacomini-White tests as described in Section II.G. The comparison is reported at the one-day horizon, which has the largest number of forecast origins.

A further check splits the out-of-sample period at July 2024 into a tightening subperiod, from May 2022 to June 2024, which covers the policy-rate increases of the Federal Reserve and Bank Indonesia, and a post-tightening subperiod, from July 2024 to July 2026. Within each subperiod, both the macroeconomic comparison and the interval comparison between the ARIMAX-GARCH model and its frozen-variance counterpart are repeated, so that the persistence of the two main findings can be assessed across different monetary conditions.

I. Computation and Reproducibility

All computations are carried out in Python 3.13 on Google Colab. NumPy 2.1 and pandas 2.2 handle data alignment, rolling-window construction, and the Engle-Granger long-run regression of the error-correction extension. The arch library (version 8.0.0) performs the unit-root tests and the joint QMLE of every GARCH-type model, and statsmodels 0.15 provides the Johansen test, the residual diagnostics, and the HAC regressions behind the Clark-West and Giacomini-White statistics. SciPy 1.16 supplies the normal and chi-square distributions used in the significance tests and in the Kupiec and Christoffersen tests. The interval simulations draw normal shocks, consistent with QMLE, from a random generator with a fixed seed of 42.

III. RESULTS AND DISCUSSION

A. Data Characteristics

The daily rupiah exchange-rate data from 2 April 2018 to 27 July 2026 comprise 2,004 trading-day observations aligned with the exogenous series. Figure 1 shows that the IDR depreciated from about IDR 13,750 per US dollar in early 2018 to a record low of about IDR 18,171 on 8 June 2026, and this persistent rise with no return to a fixed mean is a visual sign of nonstationarity. Figure 2 presents the daily log returns and their squared values, revealing pronounced volatility clustering that is most visible during the COVID-19 shock of 2020 and the tightening cycle of 2022 to 2024. These two features motivate the differencing of the level and the explicit modeling of the conditional variance.



Figure 1. Daily rupiah exchange rate, 2018 to 2026

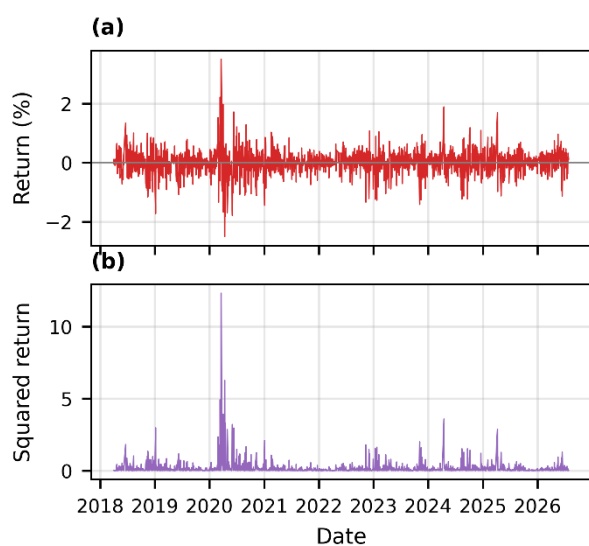


Figure 2. Daily log-returns and squared returns of the rupiah rate

B. Stationarity and Model Estimation

Table I presents the ADF and PP results. The exchange rate and the three exogenous variables fail to reject the unit-root null in levels, with p-values well above 0.05 on both tests, whereas after first differencing each series rejects the null at $p < 0.001$. The two tests agree in every case, so all series are integrated of order one, $I(1)$.

The Johansen test, with the lag order set by the AIC ($p = 2$), indicates a single cointegrating vector at the 5% level. Both the trace and the maximum-eigenvalue statistics reject the null of no cointegration, with trace = 70.14 and maximum eigenvalue = 57.03 against 5% critical values of 47.85 and 27.59, but they fail to reject the null of at most one relationship, with trace = 13.11 < 29.80 and maximum eigenvalue = 11.77 < 21.13. There is therefore a single long-run equilibrium linking the exchange rate to the macroeconomic fundamentals.

TABLE I
UNIT-ROOT TESTS FOR LEVELS AND FIRST DIFFERENCES

Series	ADF (level)	PP (level)	ADF (Δ)	PP (Δ)	Order
Log	0.6641	0.6885	0.0000	0.0000	$I(1)$
FFR	0.8531	0.8571	0.0000	0.0000	$I(1)$
BI	0.4281	0.7380	0.0000	0.0000	$I(1)$
CPI	0.9669	0.9749	0.0000	0.0000	$I(1)$

In both windows the PACF of the log returns is significant at lags 1 and 2 and cuts off at lag 3 (first window: 0.196 and 0.076; most recent window: 0.101 and -0.121; 95% band ± 0.062), which identifies an AR(2) candidate. On the first estimation window (3 April 2018 to 19 May 2022), the BIC favors AR(1)-GARCH(1,1), but no order up to AR(3) removes the residual autocorrelation (Ljung-Box $p < 0.05$ for every candidate), consistent with the COVID-19 structural break discussed in Section III.C. On the most recent window (25 May 2022 to 27 July 2026), AR(2)-GARCH(1,1) attains the lowest BIC (715.17) and yields white-noise standardized residuals (Ljung-Box $p = 0.238$), whereas AR(1) does not ($p = 0.011$). AR(2)-GARCH(1,1) is therefore adopted. The ARCH-LM test on the mean-equation residuals of the first window strongly rejects homoskedasticity at $p < 0.001$, which confirms the need for the GARCH component. The final model is an ARIMAX(2)-GARCH(1,1) with the three differenced exogenous regressors, estimated jointly by QMLE.

Table II reports the estimates in two forms. The left block comes from a representative window, the most recent 1,000 observations from 25 May 2022 to 27 July 2026, which leave 998 observations for estimation after the two autoregressive lags. The right block summarizes the estimates across all 1,004 rolling windows. In the representative window both autoregressive lags are significant, which supports the AR(2) order, and the variance-equation coefficients $\alpha_1 = 0.095$ ($p = 0.006$) and $\beta_1 = 0.860$ ($p < 0.001$) sum to $\alpha_1 + \beta_1 = 0.955 < 1$, confirming a covariance-stationary and highly persistent volatility process. The right-block summary is descriptive, because overlapping windows make the cross-window estimates dependent, so the formal test of the macroeconomic contribution rests on the nested Clark-West statistic.

TABLE II
ARIMAX(2)-GARCH(1,1) ESTIMATES, REPRESENTATIVE WINDOW AND CROSS-WINDOW DISTRIBUTION

Parameter	Coef.	Std. err.	p-value	95% CI	Mean	SD	% coef>0	% windows p<0.05
<i>Mean Equation</i>								
Constant	0.0278	0.0102	0.006	[0.008, 0.048]	0.0103	0.0096	74.3	22.0
logret(t-1)	0.1134	0.0345	0.001	[0.046, 0.181]	0.1248	0.0152	100.0	100.0
logret(t-2)	-0.1137	0.0356	0.001	[-0.183, -0.044]	-0.0657	0.0514	15.0	53.8
Δ FFR	-0.3617	0.2551	0.156	[-0.862, 0.138]	-0.2676	0.1461	4.8	0.0
Δ BI	0.3503	0.2830	0.216	[-0.204, 0.905]	0.5179	0.2068	96.2	42.2
Δ CPI	0.0892	0.1651	0.589	[-0.234, 0.413]	0.0544	0.0377	92.0	0.0
<i>Variance Equation</i>								
ω	0.0060	0.0037	0.104	[-0.001, 0.013]	0.0045	0.0016	100.0	0.6
α_1	0.0948	0.0342	0.006	[0.028, 0.162]	0.1081	0.0221	100.0	60.1
β_1	0.8600	0.0505	<0.001	[0.761, 0.959]	0.8600	0.0276	100.0	100.0

Note: Returns in percent. Representative window: 25 May 2022 to 27 July 2026, $N = 998$, log-likelihood = -326.51, AIC = 671.01, BIC = 715.17. The 95% CI is based on robust standard errors of the representative window.

C. Residual Diagnostics

The standardized residuals of the ARIMAX(2)-GARCH(1,1) model in the representative window show no remaining autocorrelation (Ljung-Box(20) = 24.09, $p = 0.238$) and no remaining ARCH effect (ARCH-LM(10) = 10.72, $p = 0.380$; Ljung-Box(20) on squared residuals, $p = 0.767$), so GARCH(1,1) captures the volatility clustering. Their mean (-0.010) and standard deviation (1.001) match the unit-variance assumption, while the Jarque-Bera test rejects normality (429.71, $p < 0.001$) because of heavy tails (kurtosis 6.21) rather than asymmetry (skewness -0.03), which QMLE with robust standard errors accommodates. The nested ARIMA-GARCH control passes the same tests (Ljung-Box 24.02, $p = 0.242$; ARCH-LM 10.17, $p = 0.426$). The residual autocorrelation of the 2018 to 2022 training period, linked to the COVID-19 break, does not affect the out-of-sample period, which begins in May 2022.

D. Point Forecast Accuracy and Significance Testing

Table III reports the point-forecast accuracy, which declines with the horizon as errors accumulate. At one day the five models are practically indistinguishable (Figure 3). From the weekly horizon onward the ARIMA-GARCH and ARIMAX-GARCH models record the lowest errors, followed closely by the RW with drift, and the macroeconomic model is even slightly worse than its control at the quarter, so the macroeconomic regressors do not improve point accuracy. The MAE reduction over the RW without drift widens to 8.7% at the quarter (Figure 4), but this is a gain in forecasting the exchange-rate level, not evidence of return predictability. It comes from the model constant, which extrapolates the depreciation trend that the driftless random walk ignores, and it largely disappears against the random walk with drift, which the model does not significantly beat beyond one day in log-return space (Table IV). The ARIMAX-ECM-GARCH records the highest error among the structured models at every horizon.

TABLE III
POINT-FORECAST ACCURACY OF THE FIVE MODELS

Horizon	Model	MAE	RMSE	MAPE
h=1	RW	41.14	55.98	0.2586
	RW-drift	40.98	55.92	0.2577
	ARIMA	40.89	55.44	0.2571
	ARIMAX	40.89	55.43	0.2571
	ARIMAX-ECM	41.12	55.65	0.2586
h=5	RW	99.13	129.93	0.6234
	RW-drift	97.53	129.39	0.6140
	ARIMA	97.46	129.05	0.6138
	ARIMAX	97.45	129.03	0.6137
	ARIMAX-ECM	100.69	131.50	0.6337
h=10	RW	148.63	189.97	0.9354
	RW-drift	145.60	188.46	0.9176
	ARIMA	144.81	187.28	0.9132
	ARIMAX	144.83	187.26	0.9133
	ARIMAX-ECM	149.78	193.16	0.9443
h=20	RW	215.73	277.59	1.3575
	RW-drift	208.89	273.08	1.3178
	ARIMA	208.36	270.92	1.3157
	ARIMAX	208.35	270.99	1.3156
	ARIMAX-ECM	216.52	282.44	1.3643
h=60	RW	419.04	501.28	2.6089
	RW-drift	383.60	471.87	2.4021
	ARIMA	381.42	471.36	2.3927
	ARIMAX	382.53	472.71	2.3995
	ARIMAX-ECM	397.57	480.89	2.4878

Note: MAE and RMSE in IDR, MAPE in %, h in trading days. RW-drift is the RW with drift, and all ARIMA-type models use GARCH(1,1) variance.

Tables IV and V report one-sided Clark-West p-values in log-return space, so their RMSE differences are not directly comparable with the level metrics in Table III. The errors of the iterated forecasts for $h \geq 5$ overlap, so the Clark-West statistic at those horizons is read together with the RMSE difference.

TABLE IV
CLARK-WEST TEST OF THE ARIMAX-GARCH MODEL AGAINST
EACH BENCHMARK

Horizon	vs RW	vs RW-drift	vs ARIMA-GARCH
h=1	0.000 (+0.97)	0.000 (+0.89)	0.066 (+0.02)
h=5	0.013 (+0.61)	0.059 (+0.24)	0.238 (+0.02)
h=10	0.014 (+1.26)	0.060 (+0.57)	0.410 (+0.01)
h=20	0.029 (+2.04)	0.143 (+0.64)	0.622 (−0.03)
h=60	0.053 (+5.11)	0.449 (−0.42)	0.899 (−0.27)

Note: One-sided Clark-West p -values in log-return space. Values in parentheses are the RMSE difference relative to the benchmark in percent, where a positive value means the model is more accurate.

TABLE V
CLARK-WEST TEST OF THE ARIMAX-ECM-GARCH MODEL
AGAINST EACH BENCHMARK

Horizon	vs RW	vs RW-drift	vs ARIMA-GARCH	vs ARIMAX-GARCH
h=1	0.000 (+0.56)	0.001 (+0.48)	0.414 (−0.39)	0.434 (−0.41)
h=5	0.299 (−1.40)	0.312 (−1.77)	0.380 (−2.00)	0.384 (−2.02)
h=10	0.246 (−2.01)	0.269 (−2.72)	0.336 (−3.30)	0.337 (−3.31)
h=20	0.171 (−2.34)	0.218 (−3.80)	0.252 (−4.50)	0.249 (−4.47)
h=60	0.068 (+2.50)	0.142 (−3.19)	0.154 (−3.03)	0.150 (−2.75)

Note: Entries and parentheses as defined in Table IV.

First, the ARIMAX-GARCH model is more accurate than the RW without drift at every horizon, but this advantage comes from the depreciation trend rather than from return predictability. The advantage is individually significant from the one-day to the monthly horizon ($p \leq 0.029$), although only the one-day gain survives the Holm adjustment across horizons, and the RMSE difference rises from 0.97% at one day to 5.11% at the quarter. The RW with drift reaches an MAE of 383.60 at the quarterly horizon against 419.04 for the RW, so the drift alone reproduces about 97% of the improvement of the full model over the no-drift baseline, without any autoregressive component or macroeconomic regressor. This confirms that the source of the advantage is the trend.

Second, the ARIMAX-GARCH model is significantly more accurate than the RW with drift only at the one-day horizon, by 0.89% in RMSE, an edge that survives the Holm adjustment (adjusted $p = 0.001$), and this edge comes from the autoregressive dynamics. The differences at the weekly to monthly horizons are small and not significant at the 5% level ($p = 0.059$ to 0.143), and the quarterly difference changes sign

(−0.42%, $p = 0.449$). The JISDOR is constructed as an average of interbank quotes, which injects first-order autocorrelation into the daily differences [47], so the one-day autoregressive edge is likely a microstructure artifact rather than exploitable predictability.

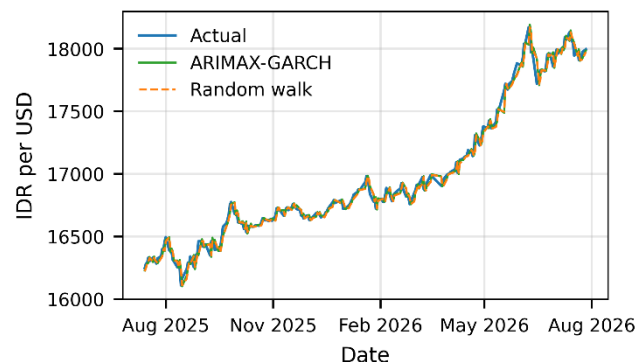


Figure 3. One-day-horizon ($h = 1$) rolling forecasts: actual vs. ARIMAX-GARCH vs. driftless random walk (last 250 days)

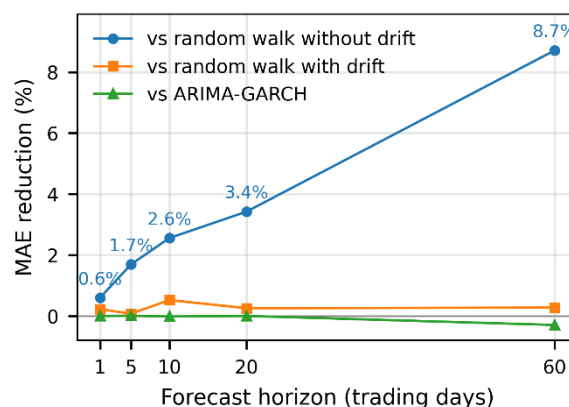


Figure 4. MAE reduction of ARIMAX-GARCH over three benchmarks by forecast horizon

Third, the comparison with the nested ARIMA-GARCH control isolates the contribution of the macroeconomic regressors. The Clark-West test does not reject equal accuracy at any horizon ($p = 0.066$ to 0.899), the RMSE gain never exceeds 0.02%, and each macroeconomic coefficient has a 95% confidence interval that contains zero (Table II). The Giacomini-White test agrees, with p -values of 0.116, 0.567, 0.847, and 0.676 at $h = 1, 5, 10$, and 20 . Its only rejection, at the quarterly horizon in favor of the nested control ($p = 0.037$), does not survive the Holm adjustment across the five horizons (adjusted $p = 0.184$), and no Clark-West comparison with the control is significant after adjustment either. The two tests therefore concur that the macroeconomic variables add no forecasting value at any horizon.

The ARIMAX-ECM-GARCH tests the long-run channel, in which a useful adjustment coefficient would pull the exchange rate back toward its equilibrium with the FFR, the BI rate, and the CPI, so that the deviation predicts the next movement. The data reject this channel. The RMSE of the ARIMAX-ECM-GARCH exceeds that of both the ARIMA-

GARCH and ARIMAX-GARCH models at every horizon, by 0.39% to 4.50%, with no significant Clark-West gain against either ($p = 0.150$ to 0.434 ; Table V), so the equilibrium deviation only adds estimation noise.

E. Interval Forecast Evaluation

Figure 5 presents the empirical coverage and the relative width of the 95% interval by horizon. At one day all models reach coverage close to the nominal level, between 93.9 and 95.5% (panel (a)), and the GARCH-based intervals are the narrowest (panel (b)), about 12% tighter than those of the two random walks and about 11% tighter than the frozen-variance ARIMAX. The conditional variance adapts to the current volatility state, which was below the full-sample average at most origins, whereas the constant-variance benchmarks fix the width at the unconditional level and cannot contract in calm periods. The GARCH models also record the lowest one-day interval score (Table VI), so this sharpness does not sacrifice coverage, and neither the Kupiec nor the Christoffersen test rejects calibration for any model at this horizon at the 5% level. The conditional-variance component therefore produces sharper and well-calibrated intervals at the one-day horizon.

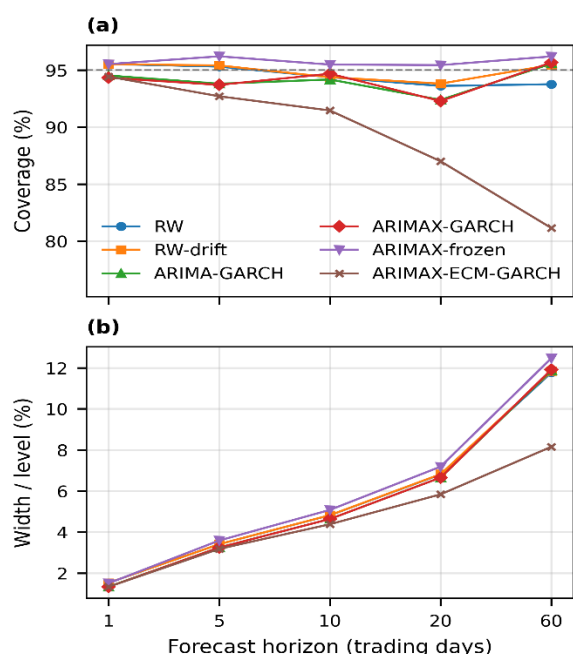


Figure 5. (a) Empirical coverage of the 95% interval by forecast horizon, with the dashed line marking the nominal 95% level. (b) Relative width of the 95% interval by forecast horizon

The macroeconomic variables add no value in the interval space either, since the ARIMAX-GARCH and ARIMA-GARCH models produce practically identical coverage and interval scores at every horizon (Table VI). The interval

advantage therefore comes from the GARCH component, and it is confined to the one-day horizon. The RW attains the lowest score at the weekly horizon, the scores are close at two weeks, and at the monthly horizon the undercoverage of the GARCH intervals (91.9 to 92.5%) lets the constant-variance intervals score clearly lower, as the conditional variance decays toward its unconditional level. The ARIMAX-ECM-GARCH narrows its spread excessively, so its coverage falls to 81.4% at the quarter. The Christoffersen test rejects independence for all models beyond one day because the multi-step forecasts overlap, so those horizons are assessed by coverage, width, and the interval score.

F. Robustness Checks

The central result is robust to the main modeling choices. The macroeconomic regressors improve the one-day RMSE by at most 0.02% in each of the six specifications (the baseline, the Student-t likelihood, windows of 750 and 1,250 observations, and EGARCH and GJR-GARCH variances), and the Giacomini-White test never rejects equal predictive ability ($p = 0.058$ to 0.433). The Clark-West statistic stays above 0.05 in five specifications ($p = 0.057$ to 0.194) and falls marginally below it only under the 1,250-observation window ($p = 0.037$), where the gain is still 0.02%, so the macroeconomic contribution remains economically negligible under every specification.

The two main findings also hold across monetary conditions. The macroeconomic regressors improve the one-day RMSE by only 0.02% in the tightening subperiod and 0.01% in the post-tightening subperiod (Clark-West $p = 0.133$ and 0.146 ; Giacomini-White $p = 0.254$ and 0.259), with no significant gain at longer horizons. The one-day GARCH interval remains 15.7% and 6.8% narrower than the frozen-variance interval in the two subperiods, with coverage of 93.7% and 94.9% and lower interval scores (296.3 against 313.2 and 290.8 against 313.0), while at the monthly horizon the frozen-variance interval scores better in both

G. Discussion and Interpretation

Figure 6 and the cross-window block of Table II show that the ΔFFR and ΔCPI coefficients are significant in none of the windows, whereas the ΔBI rate coefficient is significant in 42.2% of them with a positive sign in 96%. This in-sample association reflects a sparse step variable that changes only at policy meetings, and it does not translate into out-of-sample value, consistent with Section III.D. The null result also has a practical interpretation, namely that a forecaster who knows only past macroeconomic values, as in the design of Section II.B, cannot improve on the nested model at any horizon.

TABLE VI
COVERAGE AND INTERVAL SCORE OF THE 95% FORECAST INTERVALS BY HORIZON

Model	h=1 Cov.	h=1 IS	h=5 Cov.	h=5 IS	h=10 Cov.	h=10 IS	h=20 Cov.	h=20 IS	h=60 Cov.	h=60 IS
RW	95.5	317.5	95.3	697.4	94.4	972.5	93.6	1368.3	93.8	2284.9
RW-drift	95.5	318.1	95.4	701.7	94.4	984.9	93.8	1395.2	95.4	2216.7
ARIMA-GARCH	94.2	292.4	93.8	705.0	94.8	959.9	92.5	1474.9	95.6	2256.6
ARIMAX-GARCH	93.9	293.5	93.8	707.4	94.3	971.8	91.9	1475.2	95.8	2268.1
ARIMAX-frozen	95.3	315.0	96.1	716.5	95.2	989.1	95.5	1402.4	96.3	2299.0
ARIMAX-ECM-GARCH	94.3	293.9	93.2	703.3	91.0	1018.3	86.8	1638.3	81.4	2599.2

Note: Cov. is the empirical coverage in percent against the nominal 95%. IS is the interval score averaged across origins, where lower is better. ARIMAX-frozen fixes the variance at the unconditional GARCH variance.

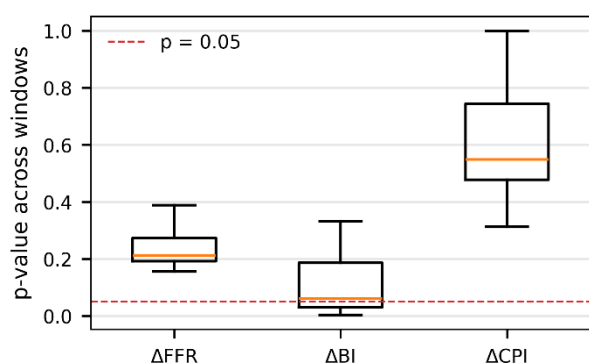


Figure 6. Distribution of the macroeconomic regressor p-values across rolling windows

The point estimates also run counter to the direction of conventional capital flows. The FFR enters with a negative sign at -0.3617 and the BI rate with a positive sign at 0.3503 . Both signs are the reverse of what a simple parity argument implies, namely that a higher US interest rate depreciates the rupiah and a higher domestic rate appreciates it. This in-sample association must not be read as predictive power, because it does not translate into out-of-sample forecast value. The exchange rate is dominated by unpredictable day-to-day noise, and the daily-frequency standard errors preclude sharp individual inference.

The two sources of the model's advantage can be stated precisely. The advantage over the RW comes from the constant of the mean equation, that is the drift, whose coefficient is positive in 74.3% of the windows and reproduces the persistent depreciation trend, so the difference accumulates with the horizon. The RW with drift carries the same component, so the two models are practically equal and both beat the RW. The thin advantage over the RW with drift, which appears only at the one-day horizon, comes from the autoregressive terms, and it signals a slight departure from weak-form market efficiency [48] at the daily frequency rather than economically meaningful predictability. Neither source involves the macroeconomic fundamentals.

These results place the rupiah literature in a different light. The rupiah studies that report a mean absolute percentage error below 2% [2][3][4][5][6][7][8], including the weekly ARIMA-GARCH model with a test MAPE of 1.99% [12], evaluate their models without a random walk with drift. The present ARIMAX-GARCH model attains an even lower one-day MAPE of 0.26%, yet its gain over the drift benchmark is economically negligible, which shows that a low MAPE alone is not evidence of predictive ability for a trending currency. The same caution applies to the very small errors reported for ARIMA, LSTM, and ensemble models without any random-walk benchmark [3][19]. The null macroeconomic contribution differs from the gains reported for the Chinese yuan with deep learning [13] and for the official Egyptian rate [16], but those studies use different model classes and exchange-rate regimes, and the second also finds no gain for the parallel-market rate. The result agrees instead with the Indonesian evidence that complex models do not significantly beat naive forecasts of inflation under rolling-origin evaluation [22] and that macroeconomic fundamentals do not predict rupiah returns outside the Asian financial crisis [30]. This study extends that evidence to daily data, to a model that accounts for conditional heteroskedasticity, and to a sample that ends in 2026.

H. Limitations

The findings are limited to three macroeconomic variables, the daily frequency, and the linear model class. The regressors are aligned to the daily grid by last-observation-carried-forward, so their first differences are zero on most trading days and monthly information is hard to detect in daily returns, which partly explains the insignificance of ΔCPI . The regressors also enter the forecasts only through their estimated coefficients, because their future differences are set to zero, and the AR(2)-GARCH(1,1) order was identified from the most recent window, which lies in the evaluation period, although the same order is imposed on both models in the central comparison. The single-equation error-correction form relies on the weak exogeneity of the three regressors with respect to the cointegrating vector, which is plausible

because all three are set or released outside the daily exchange-rate dynamics but has not been tested formally. The intervals also draw normal innovations, and although the residuals are leptokurtic, the near-nominal one-day coverage indicates that this approximation is adequate. A formal test of weak exogeneity, together with the extensions listed in Section IV, is left for future work.

IV. CONCLUSION

Macroeconomic variables do not improve daily rupiah forecasts from ARIMA-GARCH models. Adding the Federal Funds Rate, the Bank Indonesia policy rate, and the Indonesian Consumer Price Index to an ARIMA-GARCH model produced no significant gain over the nested control at any horizon under a leakage-free rolling-window protocol, according to both the Clark-West and the Giacomini-White tests. The null result held under an error-correction extension, a Student-t likelihood, estimation windows of 750 and 1,250 observations, EGARCH and GJR-GARCH variances, and two separate subperiods. This is consistent with the Meese-Rogoff puzzle for the three variables, the daily frequency, and the linear specification tested, but it does not establish that the puzzle holds for rupiah forecasting in general. The advantage of the model over the random walk without drift came from the depreciation trend, since the random walk with drift reproduced about 97% of it. The GARCH component added value in interval rather than point forecasting, giving well-calibrated 95% intervals about 12% narrower than constant-variance benchmarks at one day, an edge that faded at longer horizons.

These findings extend previous rupiah studies in two ways. They show that the low forecast errors reported without a drift benchmark cannot be read as evidence of predictive ability, and they extend the Indonesian evidence on weak macroeconomic fundamentals to daily data and to a model with conditional heteroskedasticity. Future research should test lagged macroeconomic regressors known at the forecast origin, incorporate the fundamentals at their native monthly frequency, and compare the protocol with nonlinear multivariate models such as long short-term memory networks and with a full VECM in which all variables are endogenous.

DATA AND CODE AVAILABILITY

The Python code and the datasets that reproduce every table and figure in this study are publicly available at <https://github.com/sryczu6-create/idrusd-arimax-garch>.

REFERENCES

- [1] Bank Indonesia, "Jakarta Interbank Spot Dollar Rate (JISDOR)," Bank Indonesia. Accessed: Jul. 27, 2026. [Online]. Available: <https://www.bi.go.id/id/statistik/informasi-kurs/jisdor/default.aspx>
- [2] N. A. Firdausanti, V. Forestryani, and H. M. Nuroini, "Deep learning and statistical approaches for forecasting the Indonesian Rupiah exchange rate," *Inferensi*, vol. 8, no. 2, p. 157, Jul. 2025, doi: 10.12962/j27213862.v8i2.22709.
- [3] W. A. Pratiwi, I. M. Sumertajaya, and K. A. Notodiputro, "Comparison of ARIMA, LSTM, and ensemble averaging models for short-term and long-term forecasting of non-stationary time series data," *Inferensi*, vol. 8, no. 3, pp. 231–241, Nov. 2025, doi: 10.12962/j27213862.v8i3.22643.
- [4] S. F. N. Islam, A. Sholahuddin, and A. S. Abdullah, "Extreme gradient boosting (XGBoost) method in making forecasting application and analysis of USD exchange rates against rupiah," *J. Phys. Conf. Ser.*, vol. 1722, no. 1, p. 012016, Jan. 2021, doi: 10.1088/1742-6596/1722/1/012016.
- [5] R. Arisanti, S. N. Simbolon, and R. S. Pontoh, "The extension of Moore-Penrose generalized inverse for extreme learning machine in forecasting USD/IDR exchange rate as impact of COVID-19," *Commun. Math. Biol. Neurosci.*, 2022, doi: 10.28919/cmbn/7706.
- [6] M. F. Fayyad, V. Kurniawan, M. R. Anugrah, B. H. Estanto, and T. Bilal, "Application of recurrent neural network bi-long short-term memory, gated recurrent unit and bi-gated recurrent unit for forecasting rupiah against dollar (USD) exchange rate," *Public Res. J. Eng. Data Technol. Comput. Sci.*, vol. 2, no. 1, pp. 1–10, Apr. 2024, doi: 10.57152/predatecs.v2i1.1094.
- [7] A. F. Ningrum, D. Purwanto, and A. Kusuma Wardani, "Evaluation of deep learning optimizers for predicting JISDOR exchange rates using LSTM networks," *J. Data Insights*, vol. 3, no. 1, pp. 32–41, Jun. 2025, doi: 10.26714/jodi.v3i1.726.
- [8] M. A. Hafizh, A. P. Sari, and H. E. Wahanani, "LSTM with attention optimization for IDR-USD exchange rate forecasting," *Bit-Tech*, vol. 8, no. 2, pp. 1837–1846, Dec. 2025, doi: 10.32877/bt.v8i2.3131.
- [9] P. R. Krugman, M. Obstfeld, and M. J. Melitz, *International Economics: Theory and Policy*, 10th ed. Harlow, England: Pearson, 2015.
- [10] W. Khamidah and R. Sugiharti, "Faktor yang mempengaruhi nilai tukar rupiah terhadap dolar Amerika, euro dan poundsterling," *Ecoplan*, vol. 5, no. 1, pp. 40–52, Apr. 2022, doi: 10.20527/ecoplan.v5i1.425.
- [11] Z. Z. Pratama and F. Fafurida, "Impact of United States political uncertainty on rupiah fluctuations and Indonesia's monetary indicators," *Ganaya J. Ilmu Sos. Dan Hum.*, vol. 8, no. 3, pp. 547–562, Aug. 2025, doi: 10.37329/ganaya.v8i3.4789.
- [12] N. S. Maharani *et al.*, "Aplikasi model ARIMA-GARCH dalam peramalan data nilai tukar rupiah terhadap dolar tahun 2017-2022," *J. Mat. Sains Dan Teknol.*, vol. 24, no. 1, pp. 37–50, May 2023, doi: 10.33830/jmst.v24i1.4875.2023.
- [13] S. Meng *et al.*, "Enhancing exchange rate forecasting with explainable deep learning models," in *2024 4th International Conference on Electronic Information Engineering and Computer Science (EIECS)*, Yanji, China: IEEE, Sep. 2024, pp. 892–896. doi: 10.1109/EIECS63941.2024.10800545.
- [14] T. S. Madhulatha and Md. A. S. Ghorri, "Deep neural network approach integrated with reinforcement learning for forecasting exchange rates using time series data and influential factors," *Sci. Rep.*, vol. 15, no. 1, p. 29009, Aug. 2025, doi: 10.1038/s41598-025-12516-3.
- [15] F. Türkoğlu, E. Göççek, and Y. Yumrukuz, "Predictive abilities of machine learning and deep learning approaches for exchange rate prediction," *BDDK Bankacılık Ve Finans. Piyas. Derg.*, vol. 18, no. 2, pp. 186–210, Dec. 2024, doi: 10.46520/bddkdergisi.1600294.
- [16] E. Errais and M. B. Dhafer, "Decoding Egypt's exchange rate puzzle: forecasting official and unofficial market trends with advanced models," *Int. J. Finance*, vol. 11, no. 1, pp. 36–55, Jan. 2026, doi: 10.47941/ijf.3424.
- [17] M. Fakhri and F. A. M. AlMohana, "Forecasting of the dollar exchange rate using exogenous variables with the (IV4) method and some kernel functions," *J. Econ. Adm. Sci.*, vol. 31, no. 146, pp. 156–173, Apr. 2025, doi: 10.33095/ze0mh980.
- [18] R. A. Meese and K. Rogoff, "Empirical exchange rate models of the seventies: Do they fit out of sample?," *J. Int. Econ.*, vol. 14, no. 1–2, pp. 3–24, Feb. 1983, doi: 10.1016/0022-1996(83)90017-X.
- [19] W. A. Pratiwi, I. M. Sumertajaya, and K. A. Notodiputro, "Stacking ensemble RNN-LSTM models for forecasting the IDR/USD exchange rate with nonlinear volatility," *J. Tek. Inform. Jutif*, vol. 6, no. 4, pp. 2331–2347, Aug. 2025, doi: 10.52436/1.jutif.2025.6.4.5057.
- [20] E. Nsengiyumva, J. K. Mung'atu, and C. Ruranga, "A comparative study of multivariate CNN, BiLSTM and hybrid CNN-BiLSTM

- models for forecasting foreign exchange rate using deep learning,” *Cogent Econ. Finance*, vol. 13, no. 1, p. 2526148, Dec. 2025, doi: 10.1080/23322039.2025.2526148.
- [21] M. Sahiner, “Forecasting volatility in Asian financial markets: evidence from recursive and rolling window methods,” *SN Bus. Econ.*, vol. 2, no. 10, p. 157, Sep. 2022, doi: 10.1007/s43546-022-00329-9.
- [22] R. R. Saputra, A. Wintarti, and R. W. Romadhonia, “Rolling-origin evaluation of lag-based regularized regression models for Indonesian inflation forecasting,” *CAUCHY J. Mat. Murni Dan Apl.*, vol. 11, no. 1, pp. 1381–1397, May 2026, doi: 10.18860/cauchy.v11i1.41964.
- [23] T. Gneiting and A. E. Raftery, “Strictly proper scoring rules, prediction, and estimation,” *J. Am. Stat. Assoc.*, vol. 102, no. 477, pp. 359–378, Mar. 2007, doi: 10.1198/016214506000001437.
- [24] P. F. Christoffersen, “Evaluating interval forecasts,” *Int. Econ. Rev.*, vol. 39, no. 4, p. 841, Nov. 1998, doi: 10.2307/2527341.
- [25] Federal Reserve Bank of St. Louis, “Federal Funds Target Range - Upper Limit (DFEDTARU),” FRED, Federal Reserve Economic Data. Accessed: Jul. 27, 2026. [Online]. Available: <https://fred.stlouisfed.org/series/DFEDTARU>
- [26] Bank Indonesia, “BI 7-Day Reverse Repo Rate (BI-Rate),” Bank Indonesia. Accessed: Jul. 27, 2026. [Online]. Available: <https://www.bi.go.id/en/statistik/indikator/bi-rate.aspx>
- [27] Badan Pusat Statistik, “Tingkat Inflasi Harga Konsumen Nasional Bulanan (M-to-M) (2022=100),” Badan Pusat Statistik. Accessed: Jul. 27, 2026. [Online]. Available: <https://www.bps.go.id/id/statistics-table/1/OTEzIzE=tingkat-inflasi-harga-konsumen-nasional-bulanan-m-to-m---sup-1--sup---2022-100-.html>
- [28] P. Warjiyo, *Central Bank Policy Mix: Issues, Challenges, and Policy Responses: Handbook of Central Banking Studies*. Singapore: Springer Nature, 2022.
- [29] K.-B. Tang, “The precise form of uncovered interest parity: a heterogeneous panel application in ASEAN-5 countries,” *Econ. Model.*, vol. 28, no. 1, pp. 568–573, Jan. 2011, doi: 10.1016/j.econmod.2010.06.015.
- [30] B. A. Fianto, N. Laila, R. Sukmana, and M. Madyan, “Predictors of exchange rate returns: evidence from Indonesia,” *Bull. Monet. Econ. Bank.*, vol. 23, no. 2, pp. 239–252, Jun. 2020, doi: 10.21098/bemp.v23i1.1169.
- [31] J. G. MacKinnon, “Numerical distribution functions for unit root and cointegration tests,” *J. Appl. Econom.*, vol. 11, no. 6, pp. 601–618, Nov. 1996, doi: 10.1002/(SICI)1099-1255(199611)11:6<601::AID-JAE417>3.0.CO;2-T.
- [32] S. Johansen, “Estimation and hypothesis testing of cointegration vectors in gaussian vector autoregressive models,” *Econometrica*, vol. 59, no. 6, p. 1551, Nov. 1991, doi: 10.2307/2938278.
- [33] R. F. Engle, “Autoregressive conditional heteroscedasticity with estimates of the variance of United Kingdom inflation,” *Econometrica*, vol. 50, no. 4, p. 987, Jul. 1982, doi: 10.2307/1912773.
- [34] T. Bollerslev, “Generalized autoregressive conditional heteroskedasticity,” *J. Econom.*, vol. 31, no. 3, pp. 307–327, Apr. 1986, doi: 10.1016/0304-4076(86)90063-1.
- [35] 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, Jan. 2021, doi: 10.1371/journal.pone.0245904.
- [36] R. F. Engle and C. W. J. Granger, “Co-integration and error correction: representation, estimation, and testing,” *Econometrica*, vol. 55, no. 2, p. 251, Mar. 1987, doi: 10.2307/1913236.
- [37] R. F. Engle, D. F. Hendry, and J.-F. Richard, “Exogeneity,” *Econometrica*, vol. 51, no. 2, p. 277, Mar. 1983, doi: 10.2307/1911990.
- [38] R. J. Hyndman and G. Athanasopoulos, *Forecasting: Principles and Practice*, Third edition. Melbourne, Australia: OTexts, 2021.
- [39] T. E. Clark and K. D. West, “Approximately normal tests for equal predictive accuracy in nested models,” *J. Econom.*, vol. 138, no. 1, pp. 291–311, May 2007, doi: 10.1016/j.jeconom.2006.05.023.
- [40] W. K. Newey and K. D. West, “A simple, positive semi-definite, heteroskedasticity and autocorrelation consistent covariance matrix,” *Econometrica*, vol. 55, no. 3, p. 703, May 1987, doi: 10.2307/1913610.
- [41] R. Giacomini and H. White, “Tests of conditional predictive ability,” *Econometrica*, vol. 74, no. 6, pp. 1545–1578, 2006, doi: 10.1111/j.1468-0262.2006.00718.x.
- [42] S. Holm, “A simple sequentially rejective multiple test procedure,” *Scand. J. Stat.*, vol. 6, no. 2, pp. 65–70, 1979.
- [43] P. H. Kupiec, “Techniques for verifying the accuracy of risk measurement models,” *J. Deriv.*, vol. 3, no. 2, pp. 73–84, Nov. 1995, doi: 10.3905/jod.1995.407942.
- [44] T. Bollerslev, “A conditionally heteroskedastic time series model for speculative prices and rates of return,” *Rev. Econ. Stat.*, vol. 69, no. 3, p. 542, Aug. 1987, doi: 10.2307/1925546.
- [45] D. B. Nelson, “Conditional heteroskedasticity in asset returns: a new approach,” *Econometrica*, vol. 59, no. 2, p. 347, Mar. 1991, doi: 10.2307/2938260.
- [46] L. R. Glosten, R. Jagannathan, and D. E. Runkle, “On the relation between the expected value and the volatility of the nominal excess return on stocks,” *J. Finance*, vol. 48, no. 5, pp. 1779–1801, Dec. 1993, doi: 10.1111/j.1540-6261.1993.tb05128.x.
- [47] H. Working, “Note on the correlation of first differences of averages in a random chain,” *Econometrica*, vol. 28, no. 4, p. 916, Oct. 1960, doi: 10.2307/1907574.
- [48] E. F. Fama, “Efficient capital markets: a review of theory and empirical work,” *J. Finance*, vol. 25, no. 2, p. 383, May 1970, doi: 10.2307/2325486.