

Regime-Conditioned Causal Attribution and Forecasting of the Rupiah

Jason Jesse Joel Polii¹, Puguh Hiskiawan^{2*}

^{1,2} Data Science Study Program, Universitas Bunda Mulia, Jakarta, Indonesia
s36230012@student.ubm.ac.id¹, phiskiawan@bundamulia.ac.id^{2*}

Article Info

Article history:

Received 2026-09-01

Revised 2026-10-09

Accepted 2026-10-17

Keyword:

*Causal Attribution,
Exchange-Rate Forecasting,
Regime Switching,
SHAP-FEVD Reconciliation,
Structural Decomposition.*

ABSTRACT

This study asks whether the attribution of a predictive model of the Rupiah (IDR/USD) agrees with an independent structural decomposition, separately within volatility regimes. Hybrid econometric and machine-learning studies are evaluated almost exclusively by forecast error, leaving untested whether the learning stage weights the drivers the structural stage identifies. The pipeline combines a Structural Vector Autoregression (SVAR) with Cholesky identification, four-state volatility Markov-switching labeling, regime-conditioned features, walk-forward XGBoost, and SHAP attribution reconciled per regime with the Forecast Error Variance Decomposition (FEVD), on daily data from January 2010 to June 2026. The FEVD shows own-shock dominance of 92.5 percent at the 22-day horizon, robust to alternative orderings and to identification through heteroskedasticity (82.0 percent), though sub-period estimates range from 78.8 to 95.3 percent. On direction, XGBoost attains 56.4 percent at $t+1$ (Pesaran-Timmermann, p less than 0.001); on magnitude it does not beat a random walk at any horizon, whereas a GRU baseline does, so the predictive claim is confined to direction and gradient boosting is retained for its exact tree attribution. Exact permutation tests over all relabelings of the driver axis find agreement significant in only one of twelve regime-horizon cells, the tail regime at 22 days (cosine 0.902, $p = 0.036$), carried by 2020, while the own-dynamics share rises monotonically from 18.1 to 42.4 percent across regimes, significant at every horizon. The contribution is a falsifiable per-regime attribution audit with an explicit account of where it fails.



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

I. INTRODUCTION

The stability of the Rupiah against the United States Dollar (IDR/USD) is one of the pillars of Indonesia's macroeconomic health. As an open economy with a high dependence on capital-goods imports, exchange-rate movements affect inflationary pressure, the burden of external debt, and export competitiveness [1, 2]. Over the past decade the Rupiah has shown a structural depreciation trend interspersed with episodes of sharp pressure tied to global uncertainty, so its dynamics cannot be understood adequately through a single regime.

The economic literature has identified a number of relatively well-established determinants of the Rupiah exchange rate, including the policy rate, inflation, foreign-exchange reserves, commodity prices, and external factors such as the interest-rate policy of the United States central bank and geopolitical tensions [3–5]. However, the

relationships among these variables are unstable and regime-dependent, which calls for a modeling framework capable of capturing non-linear behavior while also explaining the causal structure.

Two dominant modeling traditions confront this challenge with complementary strengths. Linear econometric models such as the Vector Autoregression (VAR) excel at structural interpretation through the Impulse Response Function (IRF) and the Forecast Error Variance Decomposition (FEVD), but are weak on non-linear predictive accuracy [6]. Machine-learning and deep-learning models such as Long Short-Term Memory (LSTM) and gradient boosting capture non-linear patterns with high accuracy [7–9], a capability demonstrated across domains as varied as medical-image segmentation, geophysical tomography, and audio recognition [10–12], but they tend to behave as black boxes. This study places causal attribution, that is, the explanation of the causal structure driving Rupiah

depreciation per regime, as its primary objective, and multi-horizon forecasting as a supporting objective.

Efforts to bridge the two traditions have been pioneered through hybrid approaches that combine econometrics with machine learning, and generally lower forecast error relative to pure models [13–16]. In parallel, awareness of regime dependence has motivated Markov-based regime-switching modeling to capture turning points induced by crises [17–20].

Read critically, however, these strands stop short of the problem posed here, and for three distinct reasons. First, the hybrid literature evaluates the merger almost exclusively by forecast error: the econometric component supplies inputs or residuals to the learning component, and success is declared when RMSE falls [13–15]. Whether the learning stage weights the same drivers the structural stage identifies is not tested, so a hybrid may gain accuracy from statistical regularities that carry no structural content. Second, the regime-switching literature establishes convincingly that exchange-rate parameters shift across latent states [17–19], but treats the regime as an estimation device rather than as a unit of attribution; these studies report which state prevails and how volatility differs across states, not how the composition of drivers is reweighted within each state. Third, the explainable-AI literature in finance now supplies mature per-observation attribution [21–23] and even time-

varying attribution through rolling windows [24], yet validates those attributions against predictive performance, ablation, or expert plausibility rather than against an independent structural benchmark. An attribution can be stable, plausible, and still structurally wrong, because a gradient-boosting model is free to load on whichever correlated proxy is most convenient.

Two gaps follow directly. First, forecast accuracy and causal attribution remain separate agendas, so accurate models rarely explain their drivers and interpretable models are rarely accurate. Second, and more specifically, no study of the Rupiah has used regime identification as the unit at which a predictive model's attribution is confronted with a structural variance decomposition, which is the only way to establish that an accurate model is accurate for structurally defensible reasons.

This regime dependence is not merely a theoretical conjecture. Figure 1 shows the rolling annualized volatility of IDR/USD over 2010–2026: long calm periods punctuated by sharp spikes during the 2013 taper tantrum, the emerging-market pressure of 2015 and 2018, the 2020 pandemic, and the global monetary tightening of 2022. It is this alternating calm-pressure structure that causes single-regime models, whether linear or non-linear, to fail to explain when and why the composition of drivers changes.



Figure 1. Rolling 22-day annualized volatility of the IDR/USD return, 2010–2026; shaded bands mark pressure episodes (2013 taper, emerging markets 2015 and 2018, 2020 pandemic, 2022 tightening).

The need for auditable causal attribution is increasingly pressing because monetary-policy decisions and exchange-rate risk management require an explanation of why a pressure episode occurs, not merely a point prediction of it. The core problem addressed by this study is how to design a framework that unifies structural causal decomposition with non-linear forecasting, and validates the prediction attribution against the structural variance decomposition consistently across regimes.

The application of SHAP to the interpretation of financial models has grown rapidly, spanning risk attribution, distress detection, and factor analysis [21–23, 25]. However, an explicit reconciliation between predictive SHAP attributions and the causal FEVD structural variance decomposition, performed separately per regime, has not been encountered in the Rupiah exchange-rate literature.

Prior work in the Indonesian context has generally focused on a single side, either econometric determinants [1] or deep-learning accuracy [7, 26]. This study positions itself to fill that accuracy-versus-attribution gap.

II. METHOD

A. Research Design

The study is designed as quantitative research of the systems-engineering type. The system built is a layered pipeline executed sequentially: (1) SVAR as the causal layer; (2) Markov-switching regime labeling and regime-conditioned econometric features as the representation layer; (3) walk-forward XGBoost as the predictive layer; and (4) SHAP reconciled with the per-regime FEVD as the interpretation layer. The design priority is per-regime causal

attribution, whereas multi-horizon forecasting serves as a test of the usefulness of the resulting feature space. Figure 2 summarizes the four sequential layers of the pipeline and the

FEVD anchor that cross-validates the predictive attribution against the structural decomposition.

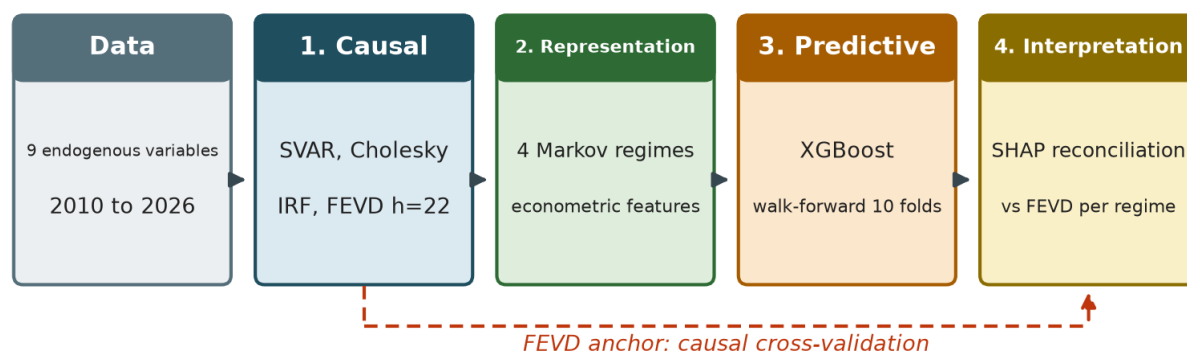


Figure 2. Research design: the four sequential layers of the pipeline (causal, representation, predictive, interpretation), with the dashed FEVD anchor linking the causal layer to the interpretation layer for per-regime cross-validation.

B. Data Collection

Data were collected into a single daily business-day master table for the period January 2010 to June 2026 (about 4,300 observations after transformation) from public sources. Nine endogenous variables are used in the causal layer: the two-year United States government bond yield (US 2Y), the VIX volatility index, an emerging-market risk proxy (EMB), the broad United States Dollar index (USD broad), Brent crude oil, gold, the Jakarta Composite Index (JCI), the BI-Fed policy-rate differential, and IDR/USD as the target. Lower-frequency series are promoted to the daily calendar through a forward-fill that carries only past values, and every interpolated cell is flagged by a marker column so that missingness information is preserved. The ADF and KPSS tests place all series at integration order one, so they are transformed into first differences (US 2Y and the BI-Fed differential) or first log-differences (the remaining series)

and then scaled. Six global variables whose markets close before the Jakarta close are shifted forward one business day (tz-safe) to prevent look-ahead information leakage across all predictive and SHAP work.

Table I summarises the resulting panel. Coverage is close to complete for every series, with forward-filled cells confined to between 0.0 and 7.1 percent of business days; the JCI carries the largest share because the Jakarta exchange observes national holidays that the global calendar does not. Eight of the nine series test as integrated of order one, so first differencing is appropriate; the VIX is the single exception, marked as requiring a check because the ADF and KPSS results on the level do not agree, and it is log-differenced with the remainder as a conservative choice. The dispersion of the level series also motivates the scaling step, since the standard deviations span roughly three orders of magnitude, from 1.566 for the US 2Y yield to 2,481.078 for IDR/USD.

TABLE I

Characteristics of the daily panel dataset (January 2010 to June 2026, business days, maximum N of 4,301 observations)

Variable (unit)	Role	Source	N	Interp. (%)	Mean	SD	Order
US 2Y (percent)	G	FRED (DGS2)	4301	4.2	1.691	1.566	I(1)
VIX (index points)	G	FRED (VIXCLS)	4301	2.9	18.416	6.771	check
EMB / EM risk (USD)	G	yfinance (EMB)	4300	3.6	105.390	10.497	I(1)
USD broad (index points)	G	FRED (DTWEXBGS)	4301	4.7	109.132	12.013	I(1)
Brent (USD/barrel)	G	FRED (DCOILBRETEU)	4301	3.1	77.998	24.212	I(1)
Gold (USD/troy oz)	G	yfinance (GC=F)	4300	3.6	1753.105	773.058	I(1)
JCI (index points)	D	yfinance (^JKSE)	4300	7.1	5584.230	1329.831	I(1)
BI-Fed differential (percent)	D	BI (manual) - FRED (DFF)	4301	0.0	4.168	2.265	I(1)
IDR/USD (IDR per USD)	T	yfinance (IDR=X)	4301	0.5	13182.934	2481.078	I(1)

Note: Role G = global, D = domestic, T = target (IDR/USD). Interp. (%) = the share of business days that are forward-filled, not original observations. Mean/SD are computed on the original level before transformation. Order refers to the combined ADF/KPSS test (Data Analysis subsection); VIX is marked *check* because ADF and KPSS on the level do not fully agree. The BI-Fed differential and US 2Y are first-differenced; the remaining series are first log-differenced.

C. Structural Layer: Identification and Decomposition

Let $y_t \in R^K$, $K = 9$, collect the stationary transforms of the nine endogenous series. The reduced-form VAR is

$$y_t = c + \sum_{i=1}^p A_i y_{t-i} + u_t, E[u_t u_t'] = \Sigma_u, \quad (1)$$

estimated with $p = 6$ selected by AIC. Structural shocks ε_t with $E[\varepsilon_t \varepsilon_t'] = I_K$ are recovered through

$$u_t = B \varepsilon_t, \Sigma_u = B B', B \text{ lower triangular}, \quad (2)$$

so that B is the Cholesky factor of Σ_u . Writing the moving-average form $y_t = \mu + \sum_{s=0}^{\infty} \Phi_s u_{t-s}$, the share of the h -step forecast-error variance of variable j attributable to structural shock k is

$$\theta_{jk,h} = \frac{\sum_{s=0}^{h-1} (e_j' \Phi_s B e_k)^2}{\sum_{m=1}^K \sum_{s=0}^{h-1} (e_j' \Phi_s B e_m)^2}, \quad (3)$$

with e_j the j -th unit vector. Because the Rupiah's own shock absorbs most of this variance, the object to be reconciled with the model attribution is the *external driver budget*, that is, the decomposition renormalized over the eight non-own shocks:

$$f_{k,h} = \frac{\theta_{jk,h}}{\sum_{m \neq j} \theta_{jm,h}}, k \neq j, \quad (4)$$

where j indexes IDR/USD. Confidence intervals for (3) come from a recursive residual bootstrap with 500 replications at the 90 percent level.

The Cholesky ordering is not arbitrary and its economic content must be stated, since (2) imposes that a shock to a variable placed later cannot affect earlier variables contemporaneously. The ordering used is US 2Y, VIX, EMB, USD broad, Brent, gold, JCI, BI-Fed differential, IDR/USD. It follows an exogeneity gradient justified by market microstructure rather than by convenience: the six global series trade in markets that close after the Jakarta session, so within the same business day they cannot respond to Indonesian price formation, whereas Indonesian assets observe them fully. Within the global block, the United States policy-rate expectation is placed first as the deepest driver of global funding conditions, followed by the global risk price (VIX), its emerging-market transmission (EMB), the Dollar numeraire, and then commodities. The two domestic non-target series precede IDR/USD because the equity index and the policy-rate differential are set in sessions that the exchange rate observes. IDR/USD is placed last, which is the conservative choice: it maximizes the variance that can be assigned to external shocks, so own-shock dominance reported under this ordering cannot be an artifact of placing the target early.

Ordering remains an assumption, so identification is stressed four ways: three alternative Cholesky orderings that permute the global block and promote IDR/USD, a sign-restriction scheme that identifies a risk-off shock by its

expected response pattern, identification through heteroskedasticity in the sense of Rigobon, which requires no ordering at all, and an SVECM comparator estimated because the Johansen diagnostic finds two cointegrating vectors. The SVECM is run both on the return object, to test whether neglected cointegration inflates own-shock dominance, and on the long-run log-level object, where the level restriction changes the economic question rather than merely the estimator.

Sample stability is tested separately. The augmented Dickey-Fuller and KPSS tests used to establish integration order assume no break, so a Zivot-Andrews test with an endogenous break is applied to both the log level and the return. Parameter constancy of the IDR/USD equation is tested with a sup-Wald (Quandt-Andrews) statistic over all candidate break dates in the central 70 percent of the sample, with the null distribution obtained by a Rademacher wild bootstrap of 499 replications because the asymptotic distribution depends on the trimming fraction, and with a CUSUM test on the OLS residuals as a complement. The decomposition (3) is then re-estimated on four economically defined sub-periods, 2010–2013, 2014–2019, 2020–2021, and 2022–2026, and on rolling 750-day windows, so that the headline share is reported as a range over the sample rather than as a single point.

D. Regime Labeling

A Markov-switching autoregression with state-dependent mean and variance is fitted to the IDR/USD return, producing volatility tiers labeled calm, base, stressed, and tail. The number of states is not assumed: models with two through five states are compared on the Bayesian information criterion, on the separation of the fitted state variances, and on whether every state retains a usable share of observations, and the selected specification is reported in the results with the criteria that support it. Regime labels enter all downstream work as *filtered* (one-sided) probabilities, which use information up to t only; the smoothed probabilities, which condition on the full sample, are used for descriptive figures alone and never for prediction or attribution.

E. Feature Construction and Real-Time Availability

The representation layer builds regime-conditioned econometric features covering trailing rolling statistics, lags, momentum, EWMA volatility, cross-series co-movement, and interactions with a real-time one-hot regime indicator. Construction is deliberately backward-looking: every window is trailing, every cross-series term uses the tz -safe lagged global block, and the regime one-hot comes from the filtered probability, so each feature is computable at the close of day t for a prediction of $t + h$. Table II lists the feature families that reach the model, their counts, their source variables, and the specific mechanism that makes each family available in real time. Conditional volatility is obtained from a GARCH(1,1) with Student- t innovations; persistence is close to one, so the Markov regime serves as

the primary discrete-state variance model and GARCH as the continuous filter. The predictive target is the volatility-scaled return, which neutralizes the ARCH effect. This feature engineering is infrastructure that supplies named, interpretable features to the attribution layer; it is not claimed as a contribution.

Two points about this inventory matter for the validity of everything downstream. First, the model receives 59 inputs in total and there is no selection stage: the 54 deterministic transforms in Table II are all constructed unconditionally, and the conditional-volatility and regime inputs are appended, so no feature is chosen by consulting the target and no selection bias enters the walk-forward. Second, the timing discipline is testable rather than asserted. A parallel variant, denoted $\text{XGB}_{\text{naive}}$, is built identically except that the global block is left unshifted, which grants the model same-day global information it could not have had at the Jakarta close. That variant is reported throughout as a timing-sensitivity upper bound and never as a result, and the gap between it and the tz-safe model measures directly how much apparent skill a look-ahead of one day would manufacture.

TABLE II

Feature families supplied to the predictive layer and the mechanism securing real-time availability

Family	Count	Availability at close of day t
Return, all nine series	9	Six global series enter lagged one business day (tz-safe)
Lags 1 and 5	18	Strictly past values
Momentum, trailing 5- and 22-day mean	18	Trailing windows ending at t
Trailing 22-day volatility (IDR/USD, USD broad, VIX)	3	Trailing window ending at t
Levels (VIX, EMB, BI-Fed differential, US 2Y)	4	Global levels lagged one business day
Trailing 22-day correlation with USD broad and EMB	2	Trailing window ending at t
GARCH(1,1) σ_t	1	Refit on fold training data, one-sided filter
Regime one-hot	4	Filtered probability, refit on fold training data
Total inputs	59	No selection stage applied

F. Predictive Layer and Evaluation

XGBoost regression (`tree_method hist`) is trained over this feature space for the $t + 1$, $t + 5$, and $t + 22$ horizons. Validation uses an expanding walk-forward with ten annual folds covering 2016 to 2025. Within each fold the Markov regime model, the GARCH filter, the feature scaler, and the predictor are refit on training data only, so no post-fold information enters either the features or the labels. Comparators are the random walk, the SVAR, XGBoost

without regime features, and a GRU as an external deep-learning baseline. Because a single metric cannot settle both questions the reviewers of this literature care about, every horizon is evaluated on magnitude (RMSE, MAE) and on direction (hit rate) simultaneously, with significance from the Diebold-Mariano test in HAC and block forms for magnitude, the Pesaran-Timmermann test for direction, and the Model Confidence Set for the joint ranking.

G. Interpretation Layer: SHAP-FEVD Reconciliation

SHAP attributions are computed with TreeExplainer on the test rows of each fold, using the model fitted on that fold's training data only. Let $\phi_{c,t}$ be the attribution of feature c at test row t , let C_m be the set of features derived from source variable m , and let T_r be the test rows labeled with regime r . The regime-conditional source attribution is

$$\bar{\phi}_m^{(r,h)} = \frac{1}{|T_r|} \sum_{t \in T_r} \sum_{c \in C_m} |\phi_{c,t}^{(h)}|. \quad (5)$$

These are split into the Rupiah's own dynamics, the eight external drivers, and non-variable context, and the driver block is renormalized onto the same simplex as (4):

$$s_k^{(r,h)} = \frac{\bar{\phi}_k^{(r,h)}}{\sum_{m \in D} \bar{\phi}_m^{(r,h)}}, D = \text{the eight drivers}. \quad (6)$$

Reconciliation compares $s^{(r,h)}$ with $f^{(h)}$ from (4) using three agreement statistics that fail differently, so that agreement is not an artifact of one of them:

$$\cos = \frac{s'f}{(s's)^{1/2}(f'f)^{1/2}},$$

$$\rho = \text{corr}(\text{rk}(s), \text{rk}(f)), L_1 = \sum_k |s_k - f_k|. \quad (7)$$

Cosine similarity between non-negative vectors is bounded well away from zero even for unrelated compositions, so a raw cosine of, say, 0.9 is not by itself evidence of alignment. Significance is therefore assessed by an exact permutation test: the null holds that the model's driver weighting bears no correspondence to the structural budget, and since there are only eight drivers, all $8! = 40,320$ permutations π of the FEVD vector are enumerated rather than sampled,

$$p = \frac{1}{|\Pi|} \sum_{\pi \in \Pi} \mathbf{1} \{ \cos(s, f \circ \pi) \geq \cos(s, f) \}, \quad (8)$$

and the same construction is applied to ρ . Uncertainty in $s^{(r,h)}$ itself is quantified by a cluster bootstrap that resamples the ten walk-forward folds with replacement (2,000 replications), which respects within-fold dependence; the calm-versus-tail difference in each driver share is tested from that distribution, and an omnibus test of whether the whole composition differs between tiers permutes tier labels within folds. All statistics are reported at $h = 1, 5$, and 22 and on test sub-samples split at 2020, so that alignment is not asserted from a single horizon or a single episode [24, 27].

H. Scope of Causal Claims

Because this study places an econometric decomposition and a machine-learning attribution side by side, the epistemic status of each must be stated precisely, and three objects that are easily conflated are kept apart throughout. The FEVD in (3) is a *structural* quantity: conditional on the identifying restriction in (2) being correct, it measures how much of the forecast-error variance of the Rupiah is generated by each structural shock. SHAP in (5) is a *predictive* quantity: it measures how much each feature moves this particular fitted model's output, it is defined relative to the model rather than to the economy, and it carries no causal content whatsoever on its own. A driver may receive a large SHAP share because it genuinely transmits shocks, or merely because it proxies for something unobserved that does.

The reconciliation therefore does not turn SHAP into a causal estimate, and no causal claim in this paper rests on SHAP. What (8) tests is a much weaker and verifiable proposition: whether the predictive model, which was never shown the structural decomposition, nevertheless reweights its drivers in the order the structural decomposition implies, and whether that correspondence strengthens or weakens by regime and horizon. Agreement is evidence that the model's accuracy is being earned on structurally defensible variation rather than on incidental correlation; disagreement flags an attribution that should not be given an economic reading. Neither direction identifies a causal effect that the SVAR has not already identified, and neither substitutes for the identifying assumption behind it. Variance shares are likewise not effect sizes: a shock can explain little variance because it was quiet over the sample and still matter greatly when it arrives.

III. RESULTS AND DISCUSSION

A. Causal Decomposition and Regime Identification

The FEVD of the IDR/USD return shows that forecast variance is dominated by own shocks at 92.5 percent at the 22-day horizon (90 percent bootstrap interval [88.8; 93.3]; 92.6 percent at $h=10$), stable across all three Cholesky orderings (Figure 3). Table III collects every identification scheme tested. The share survives the schemes that relax the ordering assumption in different directions: 92.6 percent under both alternative orderings, 93.0 percent under an SVECM on the same return object, and 82.0 percent under identification through heteroskedasticity, which imposes no ordering at all. Treating the VIX in log level rather than log difference, the one series whose integration order is ambiguous, moves the share by 0.1 point. Own-shock dominance is therefore neither an artifact of the recursive ordering nor of neglected cointegration nor of that single transformation choice.

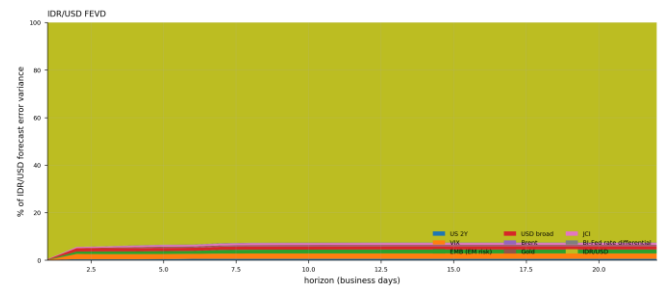


Figure 3. FEVD of the IDR/USD return (Cholesky identification): own-shock dominance of 92.5 percent at the 22-day horizon, stable across orderings.

TABLE III
IDR/USD own-shock variance share under every identification scheme tested

Scheme	Own (%)	Note
Cholesky baseline, $h=22$	92.5	[88.8; 93.3]
Cholesky baseline, $h=10$	92.6	[88.9; 93.3]
Ordering: USD first, $h=10$	92.6	Global block permuted
Ordering: commodities first, $h=10$	92.6	Global block permuted
VIX in log level, $h=22$	92.6	vs 92.5 baseline
SVECM, return object	93.0	$h=10$ and $h=22$
Heteroskedasticity (Rigobon), $h=10$	82.0	No ordering required
SVECM, log-level object, $h=10$	77.6	[69.0; 83.2]
SVECM, log-level object, $h=22$	74.2	[64.3; 81.0]

Note: Intervals are 90 percent recursive-residual bootstrap. The lower panel changes the object rather than the estimator: it decomposes the long-run log level, not the return, so it answers a different economic question. Sign-restriction identification is not an own-share scheme; it places the risk-off shock at a median 6.0 percent of return variance with a 90 percent interval of [0.8; 36.2].

Two qualifications are attached to this number rather than left implicit. First, the interval across schemes is not tight in the direction that matters for attribution: the external driver budget is 7.5 percent under the baseline but 18 percent under identification via heteroskedasticity, and the sign-restriction scheme puts the risk-off shock alone anywhere from 0.8 to 36.2 percent of return variance. The driver budget is therefore best read as roughly 7 to 18 percent, and the reconciliation below uses the conservative baseline. Second, a large own share at daily frequency is partly a construction of the identification rather than a statement that domestic fundamentals dominate. With the global block lagged one business day for timezone safety and IDR/USD ordered last, contemporaneous global information is already absorbed into the lag structure, which mechanically narrows what remains for the observed drivers. Direct evidence that excluded information exists is reported in Section III.B: the variant given same-day global data reaches materially higher directional accuracy.

Among the observed drivers the largest are the VIX (2.3 percent of total variance), emerging-market risk (1.7

percent), and USD broad (1.4 percent); renormalized onto the driver budget these become 30.4, 22.3, and 18.4 percent at $h=22$. The complementary long-run story is different in an informative way: on the log-level object the own share falls to 77.6 percent at $h=10$ (Figure 4) and the driver budget is led by USD broad, the VIX, and emerging-market risk in

near-equal proportion, which is consistent with the literature emphasizing external factors in Rupiah fluctuations [1, 4, 28]. That the level and return objects disagree is expected and is reported as such: returns are dominated by own dynamics at daily frequency, whereas the level is anchored by global conditions over long horizons.

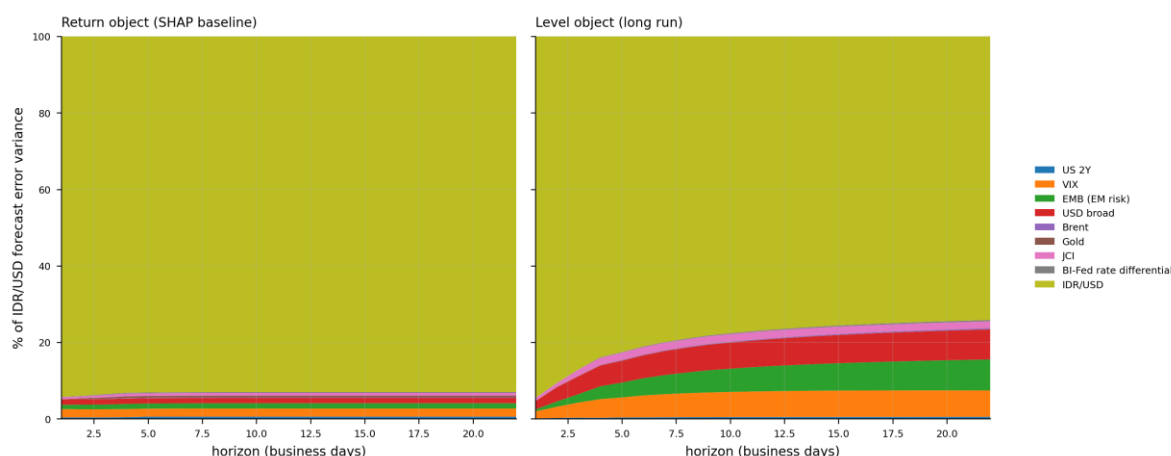


Figure 4. Dual-object SVECM FEVD: the IDR/USD own-shock share on the return object (93.0 percent) and on the long-run log-level object (77.6 percent).

Because the sample spans the taper tantrum, the pandemic, and the 2022 tightening cycle, treating the decomposition as constant over sixteen years requires evidence rather than assumption. Formal break tests find one. A Zivot-Andrews test on the log level rejects a unit root with a break at 2013-07-22 ($p=0.006$), the same test on the return breaks at 2015-09-29 ($p < 0.001$), and a sup-Wald test on the IDR/USD equation of the VAR rejects parameter constancy at 2013-11-12 (127.5, $p=0.046$), all three locating instability in the taper period. The CUSUM test on OLS residuals does not reject ($p=0.946$), so the instability is in the coefficients rather than in the residual scale.

Re-estimating the decomposition within sub-periods (Table IV) shows what this costs the headline number. Own-shock dominance is qualitatively present everywhere, but the point estimate ranges from 78.8 percent in 2014–2019 to 95.3 percent in 2010–2013, which means the external driver budget ranges from 4.7 to 21.2 percent rather than sitting at 7.5. A 750-day rolling window, stepped 25 days, gives the same message continuously: 142 windows with own shares from 64.6 to 94.2 percent, median 83.5, and 78 of the 142 below 85 percent (Figure 5). The full-sample 92.5 percent is therefore an average over a varying quantity and sits near the upper end of the rolling distribution.

TABLE IV

Sub-period re-estimation of the IDR/USD FEVD at $h=22$ (driver shares renormalized on the driver budget)

Sub-period	n	Own %	VIX	EM B	US 2Y
Full sample, $P=6$	429	92.5	30.4	22.3	7.4
	9				
Full sample, $P=2$	429	93.4	34.2	22.5	6.5
	9				
2010–2013 taper	104	95.3	8.5	21.3	4.7
	1				
2014–2019 pre-pandemic	156	78.8	29.4	34.7	2.2
	5				
2020–2021 pandemic	523	85.3	28.6	23.1	7.0
2022–2026 tightening	117	81.6	37.6	9.1	21.6
	0				

Note: Lag order is reselected within each sub-period. The $P=2$ row shows that the full-sample result is not an artifact of the lag order. Driver columns are percentages of the driver budget, not of total variance.

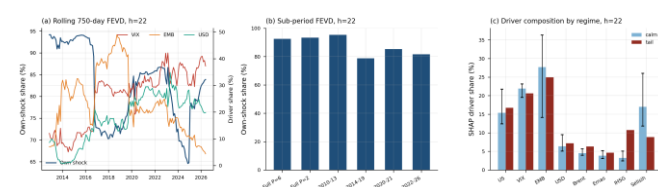


Figure 5. Structural stability. Panel (a): own-shock share and three global driver shares on a 750-day rolling window. Panel (b): sub-period own-shock shares against the full-sample estimate.

The composition of the driver budget is even less stable than its size: the VIX share moves from 8.5 percent in 2010–2013 to 37.6 percent in 2022–2026, and the US 2-year share

from 2.2 to 21.6 percent between the pre-pandemic and tightening periods, which is economically sensible given that the later period is one of policy-rate divergence. The consequence for this study is stated plainly. Conclusions about the ranking of drivers hold for the full sample and should not be read as time-invariant; the reconciliation in Section III.C is correspondingly evaluated against the full-sample budget and then re-tested on sub-samples.

The number of regimes is a modeling choice with direct consequences for the attribution below, so the evidence for it is reported in full rather than asserted. Table V compares two through five states on the same return series. The Bayesian information criterion declines monotonically across the range examined and therefore does not by itself identify a minimum; what it does show is where the marginal return collapses. Moving from two to three states buys 522 BIC points and from three to four a further 143, but the fifth state buys only 25. Four states thus sit at the elbow. Two further considerations settle the choice in the same direction: four is the smallest number that isolates an extreme-volatility state distinct from ordinary stress, which is the state the attribution analysis is about, and the four-state specification converged in all ten walk-forward folds when

refit on training data alone, which a specification used inside a walk-forward must do.

TABLE V
Selection of the number of Markov regimes on the IDR/USD return

States	Log-lik.	Par.	AIC	BIC	Δ BIC	Conv.
2	-3697.4	6	7406.8	7445.0	—	yes
3	-3411.1	12	6846.3	6922.7	522.3	yes
4	-3306.1	20	6652.3	6779.6	143.1	yes
5	-3252.0	30	6563.9	6754.9	24.7	yes

Note: Δ BIC is the improvement over the preceding row. The specification switches both mean and variance across states. Convergence here is on the full sample; the selected four-state model additionally converged in 10 of 10 walk-forward folds refit on training data only.

The selected specification yields a dominant base regime (frequency 0.60), flanked by calm (0.24) and stressed (0.16), together with an extreme tail (0.007) whose full-sample days concentrate on the 2013 taper episode and the 2020 pandemic (Figure 6). Annualized volatility climbs from 8.6 percent (calm) to 63.5 percent (tail), cross-validating those tiers against the GARCH conditional volatility.

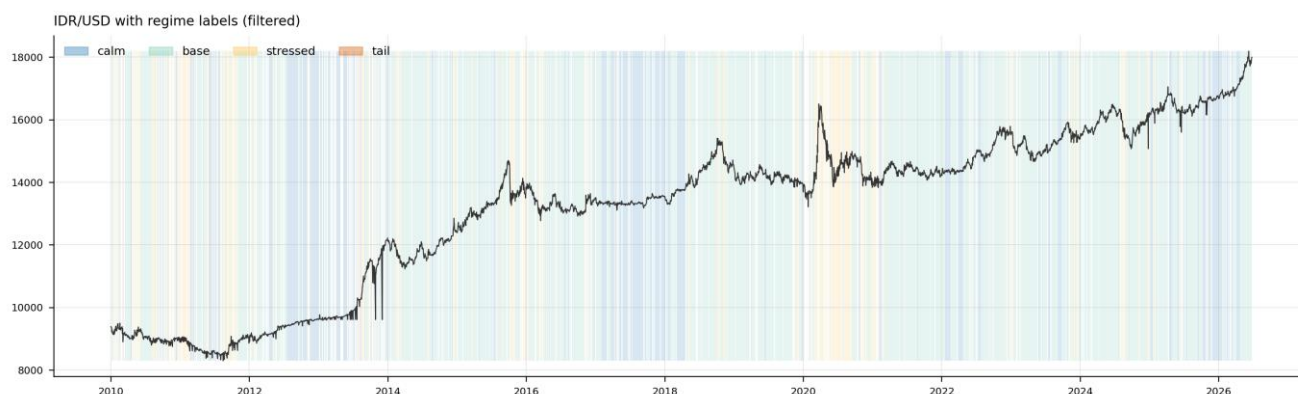


Figure 6. Four-tier volatility Markov-switching regimes over the IDR/USD return; the tail tier is concentrated on the 2013 taper episode and the 2020 pandemic.

The impulse responses of IDR/USD to the structural shocks (Figure 7) are consistent with this decomposition: the external shocks with the widest confidence bands away from zero are the VIX, emerging-market risk, and the broad

Dollar index, whereas the response to the Rupiah's own shock is the largest and dies out within a few days, matching the own-shock dominance in the FEVD.

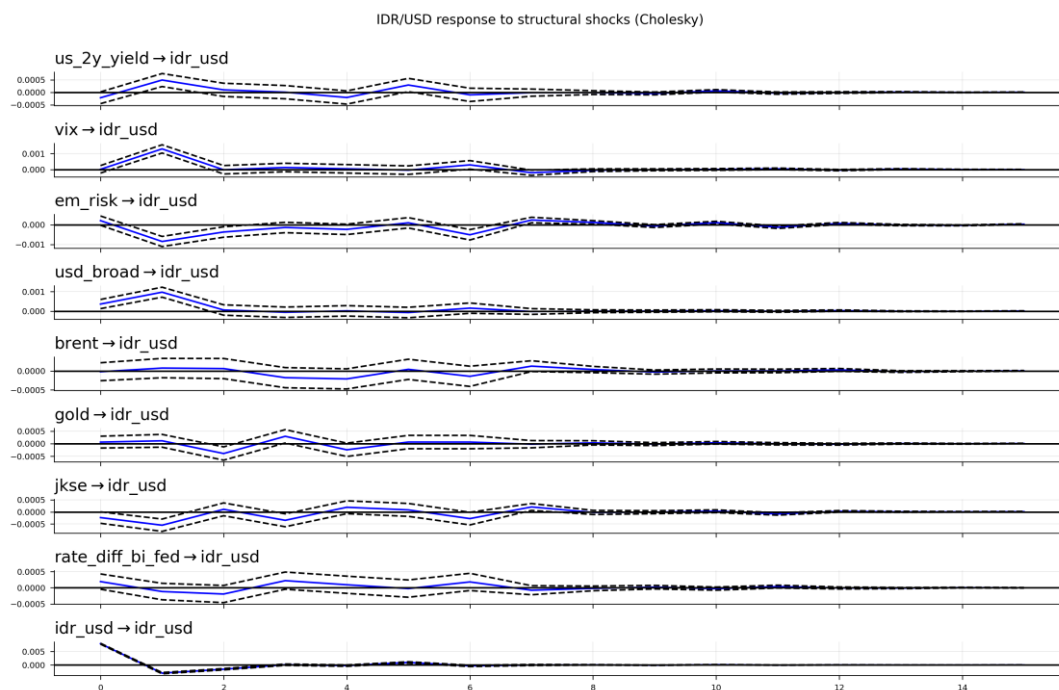


Figure 7. Impulse responses of IDR/USD to the nine structural shocks (Cholesky identification) with bootstrap confidence bands; the own shock dominates and decays within a few days.

B. Forecasting Performance

Forecast performance is reported on both dimensions at all three horizons, for all six models, with the significance

test appropriate to each dimension, because a claim restricted to one metric at one horizon cannot be assessed by a reader (Table VI). The picture that emerges is mixed and is reported as such.

TABLE VI
Walk-forward forecast performance, ten annual folds (2016–2025), $n=2,609$ test days per horizon

Model	RMSE	Hit %	DM _H	DM _B	PT p	MCS
<i>Panel A. $h = 1$</i>						
XGBoost	0.807	56.4	2.59*	1.01	<0.001	in
XGBoost, no regime	0.892	56.8	2.90*	0.99	<0.001	in
SVAR	0.651	56.2	−0.23	−0.23	<0.001	out
GRU	0.620	57.8	−3.05*	−2.42*	<0.001	in
Random walk	0.655	50.2	—	—	—	out
XGBoost _{naive}	0.789	63.7	1.50	0.79	<0.001	in
<i>Panel B. $h = 5$</i>						
XGBoost	1.445	54.3	3.82*	1.73	<0.001	in
XGBoost, no regime	1.445	53.8	3.93*	1.69	<0.001	in
SVAR	1.071	56.4	0.97	0.76	<0.001	in
GRU	1.043	56.7	−0.83	−1.00	<0.001	in
Random walk	1.055	53.5	—	—	—	in
XGBoost _{naive}	1.208	57.0	2.53*	1.18	<0.001	in
<i>Panel C. $h = 22$</i>						
XGBoost	2.843	53.8	4.34*	2.88*	0.006	out
XGBoost, no regime	2.764	54.4	4.29*	2.71*	0.001	out
SVAR	2.305	56.8	0.66	0.78	<0.001	in
GRU	2.343	56.6	1.21	2.36*	<0.001	in
Random walk	2.284	55.1	—	—	—	in
XGBoost _{naive}	2.717	53.7	4.79*	2.90*	0.018	out

Note: RMSE is in volatility-scaled return units. DM_H is the Diebold-Mariano statistic against the random walk with HAC variance, DM_B its per-fold block variant, which is the reference for $h > 1$ because horizons overlap; a positive value means worse than the random walk. * marks $p < 0.05$. PT is the Pesaran-Timmermann directional test. MCS is membership in the Model Confidence Set at size 0.10. XGBoost_{naive} is the timing-sensitivity variant with unshifted same-day global data; it is an upper bound on what a look-ahead of one day would yield, not a result.

On magnitude, the machine-learning layer does not beat the random walk at any horizon. XGBoost is worse by the HAC-variance Diebold-Mariano test at all three (t+1 RMSE 0.807 vs 0.655, $p=0.010$), although under the block variant, which is the appropriate reference where forecast horizons overlap, the difference is not significant at $h=1$ or $h=5$ and becomes significant against XGBoost only at $h=22$. What the table also shows, and what a claim confined to XGBoost would have hidden, is that two comparators do beat the random walk: the GRU is significantly better on magnitude at $h=1$ by both variants (RMSE 0.620 vs 0.655, block $p=0.039$), and the SVAR matches it at every horizon. At $h=22$ the Model Confidence Set retains only the GRU, the SVAR, and the random walk, eliminating every gradient-boosting variant. The honest summary is therefore that the gradient-boosting layer is not the best magnitude forecaster in this comparison and is not claimed to be; it is retained because tree-based attribution through exact TreeSHAP is what the reconciliation in Section III.C requires, and that is a methodological reason rather than a performance one.

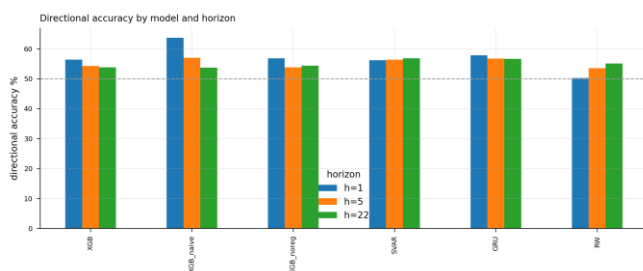


Figure 8. Forecast skill per horizon: XGBoost leads on t+1 direction but loses on RMSE magnitude to the random walk.

On direction the ranking differs. At t+1 XGBoost reaches a hit rate of 56.4 percent against the random walk's 50.2 percent, significant by Pesaran-Timmermann ($p < 0.001$), which is the one predictive claim this study makes. The edge persists but narrows with the horizon, to 54.3 percent at $h=5$ and 53.8 percent at $h=22$ ($p=0.006$), while the random walk's own hit rate rises with the horizon because of the drift in a depreciating series, so the gap closes from 6.2 points to well under two. Direction is thus forecastable at short horizons and only weakly so at monthly ones. Figure 8 sets the two dimensions side by side across horizons, the directional edge at t+1 against the magnitude shortfall relative to the random walk.

Two results bear directly on interpretation. The ablation without regime features is comparable on direction at every horizon, differing by at most 0.6 points and in both directions (56.8, 53.8, 54.4 percent against 56.4, 54.3, 53.8), so the regime features do not buy accuracy; their value is that they make attribution regime-specific, which is exactly

the claim made for them in Section II.D and not more. The timing variant is more pointed: granting the model same-day global data lifts t+1 directional accuracy from 56.4 to 63.7 percent. That 7.3-point gap is the size of the look-ahead that a careless treatment of the Jakarta close would produce, and it is larger than the entire genuine edge over the random walk. Reported results use the tz-safe construction throughout.

Directional accuracy rises monotonically across volatility tiers, namely 52.5 percent (calm), 55.9 percent (base), 61.1 percent (stressed), and 62.2 percent (tail) (Figure 9). The tail figure must not be over-read: it rests on 45 test days, of which 40 (89 percent) fall in the 2020 pandemic and none fall in 2013, since the walk-forward test period begins in 2016. It is a description of a handful of extreme episodes, not evidence of a cross-time advantage, and its Wilson interval of [48; 75] overlaps every other tier.

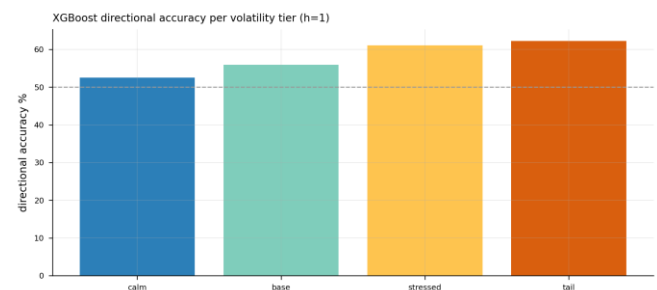


Figure 9. XGBoost t+1 directional accuracy per volatility tier with 95 percent Wilson intervals; monotonically increasing from calm to tail, with the tail resting on 45 test days.

C. Per-Regime SHAP-FEVD Reconciliation

This reconciliation is the main contribution of the study and is presented in Figure 10. The attribution split shows that the mean absolute SHAP share of the IDR/USD own dynamics rises monotonically at t+1 from 18.1 percent in the calm tier (90 percent fold-clustered bootstrap interval [17.2; 20.1]) to 23.8 [22.2; 25.1] in base, 30.8 [27.8; 34.6] in stressed, and 42.4 [41.5; 55.9] in tail, while the external driver share falls from 74.4 to 50.9 percent. The tail-minus-calm difference is 27.5 points ([21.8; 38.2], $p < 0.001$) and the same ordering holds at $h=5$ and $h=22$, where the difference is 19.4 and 18.6 points respectively, also at $p < 0.001$. This is the one regime-conditional result in the study that is significant at every horizon: as volatility rises, the model relies increasingly on the Rupiah's own dynamics, which is consistent with the high FEVD own-shock share and with the higher directional accuracy observed in the stressed and tail tiers.

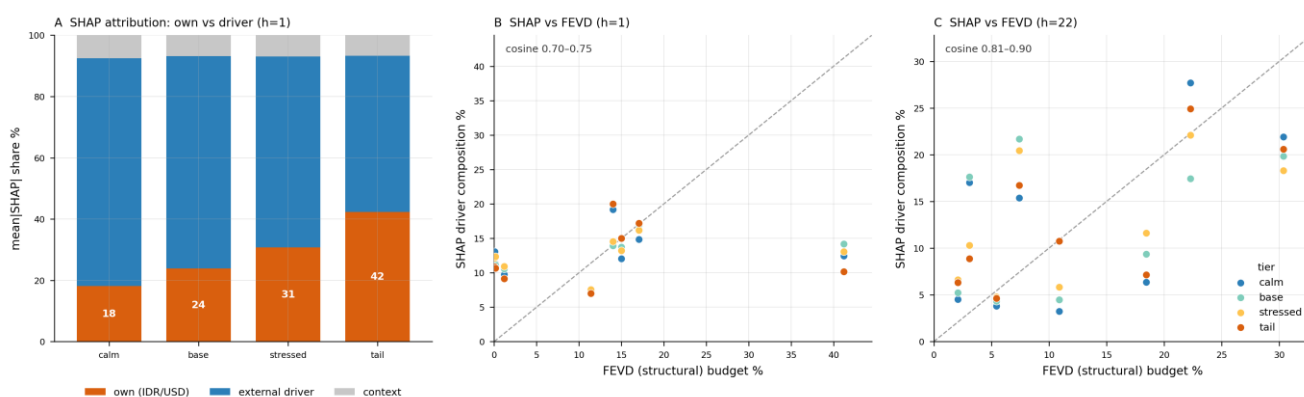


Figure 10. Per-regime SHAP-FEVD reconciliation (main contribution). Panel A: the SHAP attribution split into own-dynamics, driver, and context shares per tier, with the own share rising monotonically from calm to tail. Panel B: scatter of the SHAP driver composition per tier against the structural FEVD budget at $h=1$, loose off the diagonal owing to the Cholesky ordering's concentration on the Dollar index. Panel C: the same scatter at $h=22$, tightening to the diagonal (cosine 0.81 to 0.90) as the main reconciliation evidence.

The alignment of driver composition is horizon-dependent, and because a cosine similarity between two eight-element non-negative vectors is high even under random assignment, every value is accompanied by an exact permutation test over all $8!$ relabelings of the driver axis (Table VII). The null cosine median is between 0.65 and 0.71 throughout, which sets the scale against which the observed values must be read.

TABLE VII

Agreement between per-regime SHAP driver composition and the structural FEVD budget, with exact permutation tests

h	Tier	Cosine	p	Spearman	p	Null cos.
1	calm	0.713	0.291	0.31	0.231	0.674
1	base	0.753	0.123	0.71	0.023	0.688
1	stressed	0.726	0.249	0.62	0.058	0.693
1	tail	0.697	0.347	0.24	0.268	0.655
5	calm	0.833	0.107	0.33	0.195	0.637
5	base	0.835	0.121	0.21	0.291	0.685
5	stressed	0.793	0.197	0.38	0.163	0.681
5	tail	0.757	0.243	0.38	0.163	0.647
22	calm	0.852	0.076	0.48	0.108	0.647
22	base	0.814	0.153	0.31	0.231	0.696
22	stressed	0.876	0.055	0.57	0.066	0.705
22	tail	0.902	0.036	0.71	0.023	0.690

Note: p is the one-sided exact permutation p -value over all $8! = 40,320$ relabelings of the driver axis. Null cos. is the median cosine under that null. * marks $p < 0.05$.

Read against that null, the evidence is specific rather than global, and weaker than the raw similarity figures suggest. Of the twelve tier-horizon cells only one is significant at the 5 percent level on cosine similarity, namely the tail tier at $h=22$ (0.902, $p=0.036$), which is also significant on rank correlation (0.71, $p=0.023$). The stressed tier at $h=22$ is

borderline ($p=0.055$) as is calm ($p=0.076$). At $t+1$ no cosine is significant, which is expected since the $t+1$ FEVD is heavily concentrated on USD broad (about 41 percent of the driver budget) owing to the Cholesky ordering, leaving little composition to compare; the single significant statistic there is the base-tier rank correlation. At $h=5$ nothing is significant on either statistic.

The claim these data support is therefore narrow, and it is stated narrowly: at the monthly horizon the model's driver weighting in the high-pressure regimes moves toward the structural budget by more than relabeling chance produces, while at daily and weekly horizons, and in calm conditions, the apparent agreement is not distinguishable from chance. Sub-sample re-estimation sharpens this further and supplies the study's most important caveat. Splitting the test period, the $h=22$ tail result holds on 2020–2026 (cosine 0.900, $p=0.038$) but does not survive removing 2020 altogether (0.742, $p=0.269$, and only 5 tail test days remain). The one significant reconciliation cell is thus carried by the pandemic episode. The borderline calm and stressed cells at $h=22$ are by contrast stable across all three sub-samples (p between 0.054 and 0.103), so it is the ordering of agreement with the horizon, not the tail-tier point estimate, that is the durable part of the finding.

A complementary omnibus test asks whether the driver composition itself differs between the calm and tail tiers, permuting tier labels within folds. It does not: the observed L_1 distance is 12.2 points at $h=1$ against a null median of 11.3 ($p=0.229$) and 24.6 against 27.7 at $h=22$ ($p=0.985$). Fold-clustered bootstrap on the individual driver shares agrees, with no driver showing a significant tail-minus-calm difference at $h=1$ or $h=22$. What does differ significantly across tiers is the own-dynamics share, at $p < 0.001$ for all three horizons. The defensible regime-conditional statement is accordingly about how much the model leans on the Rupiah's own dynamics under stress, not about which external driver it substitutes toward. The reconciliation object is the relative composition of drivers, not their level; the level difference is by construction because the FEVD

decomposes the unconditional variance with own-shock dominance, whereas SHAP attributes a scaled-return model that includes the Rupiah's own lags. The consistency between the model-agnostic predictive attribution and the per-regime structural decomposition provides a basis for auditable causal attribution, in line with the direction of the explainable-AI literature in finance [25, 29–31].

D. Discussion

The single novelty claimed is the per-regime SHAP-FEVD reconciliation: validating a predictive model's attribution against an independent structural variance decomposition separately within volatility states and across horizons. The regime-conditioned feature space and the fusion of econometric features into one predictive space are infrastructure, not novelty claims. What distinguishes this design from the hybrid literature is that it can fail. A study that merges econometrics with machine learning and reports a lower RMSE cannot be contradicted by its own evidence; an attribution audit with a permutation null can be, and here it partly is. Reporting where it fails is what makes the cases where it holds informative.

Three results define what the audit establishes. The structural layer gives own-shock dominance that survives every identification scheme tried but not sub-period re-estimation, so the driver budget is a full-sample average of a quantity that moves between roughly 5 and 21 percent. The predictive layer earns a modest but significant directional edge at $t+1$ and no magnitude advantage at all, with a GRU outperforming it on magnitude, so the gradient-boosting model is justified by its exact attribution properties rather than by accuracy. The interpretation layer finds that the model's reliance on the Rupiah's own dynamics rises significantly and monotonically with volatility, while its external driver composition does not differ significantly between calm and tail states and matches the structural budget beyond chance only at the monthly horizon under stress, a result that is itself carried by 2020.

The practical implication follows from the gap between those last two findings. For a policymaker or a risk manager the question is which features carry structural content and which merely improve prediction, and the audit separates them. Features whose SHAP weight tracks the structural budget, chiefly the VIX and emerging-market risk at the monthly horizon, can be read as transmission channels and used to reason about why pressure is building. Features whose SHAP weight is high but whose agreement with the structural decomposition is indistinguishable from chance, which at the daily horizon is most of them, should be treated as predictive instruments only, useful in a hedging rule but not as evidence about mechanism. The reconciliation is therefore best used as a filter on which parts of a model's explanation may enter a policy narrative, and the appropriate operating point on the present evidence is the monthly horizon in high-volatility states.

The limitations are specific rather than generic. Identification rests on a recursive ordering, and although the own share is stable across the alternatives tested, the driver

budget is not: it doubles under identification through heteroskedasticity and the sign-restriction interval is wide. The tail regime contains 45 test days, 89 percent of them from the 2020 pandemic, which makes every tail-specific figure descriptive; the one significant reconciliation cell does not survive removing that year. The decomposition is not stable over sub-periods. The daily frequency and the timezone-safe lag mechanically inflate the own share, and the same-day variant shows that the excluded global information is worth 7.3 points of directional accuracy, so part of what the structural layer assigns to own shocks is intraday global transmission that this design cannot recover. Finally, the study uses nine endogenous public series and no intervention, capital-flow, or order-flow data, and the attribution is on mean absolute SHAP magnitude, which carries no sign.

IV. CONCLUSION

This study designed and tested a layered pipeline for the regime-based attribution and forecasting of the Rupiah exchange rate, with attribution as the priority. Structural decomposition gives own-shock dominance of 92.5 percent at $h=22$, stable across two alternative orderings, an SVECM, and identification through heteroskedasticity, but ranging from 78.8 to 95.3 percent across sub-periods; the VIX and emerging-market risk are the largest external drivers in the full sample. Four volatility states are selected on the elbow of the BIC path together with per-fold convergence. On forecasting, the model attains a significant $t+1$ directional accuracy of 56.4 percent and no magnitude advantage over a random walk at any horizon, while a GRU baseline does beat the random walk on magnitude at $t+1$; the predictive claim is therefore confined to direction, and the gradient-boosting layer is retained for exact tree attribution.

The reconciliation yields one robust and one fragile finding, and both are reported. Robust: the attribution share of the Rupiah's own dynamics rises monotonically from 18.1 to 42.4 percent from the calm to the tail state, a difference significant at $p < 0.001$ at every horizon. Fragile: agreement between the SHAP driver composition and the structural budget exceeds its permutation null in only one of twelve tier-horizon cells, the tail state at $h=22$, and that cell is carried by the 2020 episode. The contribution is thus a falsifiable per-regime attribution audit together with an honest account of the conditions, monthly horizon and elevated volatility, under which it currently holds.

Future research is advised to test structural identification beyond the Cholesky ordering more thoroughly, for example expanding sign restrictions and heteroskedasticity-based identification across all horizons, in order to free the conclusions further from the ordering assumption. The SHAP-FEVD reconciliation can be extended to signed directional attribution, not just the mean absolute magnitude, so that the direction of each driver's influence can be compared directly with the impulse response. The sample imbalance in the tail tier suggests adding crisis episodes or an approach that stabilizes the estimation of extreme

regimes. Additional competitive baselines such as ARIMA-GARCH and fuzzy-Markov time-series models could strengthen the forecast comparison. Finally, integrating the conditional-volatility output into decision-support artifacts is recommended to reinforce policy relevance.

REFERENCES

- [1] C. P. Silalahi *et al.*, "Analysis of determinants of the rupiah exchange rate against the US dollar in 2016:Q1–2023:Q4," *Formosa Journal of Applied Sciences*, 2024. <https://doi.org/10.55927/fjas.v3i8.10645>.
- [2] E. Pratiwik *et al.*, "Rupiah exchange rate: The determinants and impact of shocks on the economy," *Jurnal Ekonomi dan Studi Pembangunan*, 2023. <https://doi.org/10.18196/jesp.v24i1.18016>.
- [3] Rakhmat *et al.*, "The influence of exchange rate and foreign capital on the performance of inflation targeting framework in Indonesia," *International Journal of Sustainable Development and Planning*, 2022. <https://doi.org/10.18280/ijstdp.170523>.
- [4] R. Fredo *et al.*, "Analysis of the exchange rate level of emerging market countries during the COVID-19 outbreak," *Jurnal Ilmiah Manajemen, Ekonomi dan Akuntansi (MEA)*, 2024. <https://doi.org/10.31955/mea.v8i1.3795>.
- [5] T. Sodnomdavaa *et al.*, "Forecasting central bank policy rates using machine learning and deep learning approaches," *International Journal of Advanced and Applied Sciences*, 2026. <https://doi.org/10.21833/ijaas.2026.01.025>.
- [6] D. Hana *et al.*, "Multivariate time series analysis of the USD/IQD exchange rate using VAR, SVAR, and SVECM models," *Statistics, Optimization and Information Computing*, 2025. <https://doi.org/10.19139/soic-2310-5070-2371>.
- [7] M. Z. Abedin *et al.*, "Deep learning-based exchange rate prediction during the COVID-19 pandemic," *Annals of Operations Research*, 2021. <https://doi.org/10.1007/s10479-021-04420-6>.
- [8] J. Gu *et al.*, "AB-LSTM-GRU: A novel ensemble composite deep neural network model for exchange rate forecasting," *Computational Economics*, 2024. <https://doi.org/10.1007/s10614-024-10754-7>.
- [9] H. Huang *et al.*, "Exchange rates forecasting and trend analysis from machine learning," *Academic Conferences Series*, 2024. <https://doi.org/10.62381/acs.sdit2024.19>.
- [10] P. Hiskiawan, M. K. Sari, R. E. Siregar, N. A. Valencia, and T. P. Wijayanti, "A deep learning data fusion approach for prostate MRI zonal segmentation using dual-channel U-Net with T2 and ADC images," in *2026 International Conference on Current Research in Artificial Intelligence and Data Science (ICCRAIDS)*, (Banyumas, Indonesia), pp. 1–7, 2026. <https://doi.org/10.1109/ICCRAIDS67816.2026.11519671>.
- [11] P. Hiskiawan, C. C. Chen, and Z. K. Ye, "Processing of electrical resistivity tomography data using convolutional neural network in ERT-NET architectures," *Arabian Journal of Geosciences*, vol. 16, p. 581, 2023. <https://doi.org/10.1007/s12517-023-11690-w>.
- [12] P. Hiskiawan, T. P. Wijayanti, H. Heryanto, S. Everlin, S. A. Yasodhara, and D. Alexander, "Mel-frequency cepstral coefficients and neural networks for Indonesian traditional music recognition," in *2025 International Conference on Informatics, Multimedia, Cyber and Information System (ICIMCIS)*, (Jakarta, Indonesia), pp. 1707–1712, 2025. <https://doi.org/10.1109/ICIMCIS68501.2025.11327436>.
- [13] R. N. Okafor *et al.*, "Enhancing vector autoregression with machine learning techniques for causal analysis of macroeconomic indicators in African economies," *Asian Journal of Probability and Statistics*, 2026. <https://doi.org/10.9734/ajpas/2026/v28i2868>.
- [14] H. İ. Erçen *et al.*, "The impact of macroeconomic sustainability on exchange rate: Hybrid machine-learning approach," *Sustainability*, 2022. <https://doi.org/10.3390/su14095357>.
- [15] M.-L. Shen *et al.*, "An effective hybrid approach for forecasting currency exchange rates," *Sustainability*, 2021. <https://doi.org/10.3390/su13052761>.
- [16] R. Kler *et al.*, "Enhancing short run exchange rate forecasting for Uzbeki soum: From econometric models to machine learning-based hybrid approaches," in *2025 International Conference on Technology Enabled Economic Changes (InTech)*, 2025. <https://doi.org/10.1109/intech64186.2025.11198545>.
- [17] C. Walid *et al.*, "Stock market volatility and exchange rates in emerging countries: A Markov-state switching approach," *Emerging Markets Review*, 2011. <https://doi.org/10.1016/j.ememar.2011.04.003>.
- [18] S. Das *et al.*, "Predicting regime switching in BRICS currency volatility: A Markov switching autoregressive approach," *DECISION*, 2021. <https://doi.org/10.1007/s40622-021-00275-9>.
- [19] C. A. G. Silva, "The regime changing behavior of exchange rates and stock market prices of selected emerging countries: An application of the Markov switching vector autoregressive model (MS-VAR)," *American Journal of Economics and Business Innovation*, 2023. <https://doi.org/10.54536/ajebi.v2i1.993>.
- [20] L. Kechim *et al.*, "Estimating currency crisis probabilities in MENA countries using Markov switching models," *International Journal of Research in Economics and Finance*, 2026. <https://doi.org/10.71420/ijref.v3i1.219>.
- [21] K. L. Tran *et al.*, "Explainable machine learning for financial distress prediction: Evidence from Vietnam," *Data*, 2022. <https://doi.org/10.3390/data7110160>.
- [22] Y. Wang *et al.*, "Interpretable machine learning framework for corporate financialization prediction: A SHAP-based analysis of high-dimensional data," *Mathematics*, 2025. <https://doi.org/10.3390/math13152526>.
- [23] J. E. Choi *et al.*, "Vector SHAP values for machine learning time series forecasting," *Journal of Forecasting*, 2024. <https://doi.org/10.1002/for.3220>.
- [24] A. Ben Mrad *et al.*, "Interpretable machine learning for price index forecasting: A case study with rolling windows and SHAP," *Engineering, Technology and Applied Science Research*, 2026. <https://doi.org/10.48084/etasr.14863>.
- [25] D. S. Kumar, "Explainable AI in financial forecasting using time series analysis," *International Journal for Research in Applied Science and Engineering Technology*, 2025. <https://doi.org/10.22214/ijraset.2025.70080>.
- [26] F. E. Sönmez *et al.*, "Exchange rate prediction under data volatility: A comparison of deep learning techniques," *Journal of Metaverse*, 2026. <https://doi.org/10.57019/jmv.1841380>.
- [27] A. V. Ponce-Bobadilla *et al.*, "Practical guide to SHAP analysis: Explaining supervised machine learning model predictions in drug development," *Clinical and Translational Science*, 2024. <https://doi.org/10.1111/cts.70056>.
- [28] X. Wu *et al.*, "Analysis of the transmission mechanism of exchange rate fluctuations in emerging markets to global portfolio risk based on deep learning," *Journal of Economics, Business and Management*, 2025. <https://doi.org/10.18178/joebm.2025.13.2.863>.
- [29] 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)*, pp. 892–896, IEEE, 2024. <https://ieeexplore.ieee.org/document/10800545/>.
- [30] V. Govindaraj *et al.*, "Explainable transformers in financial forecasting," *World Journal of Advanced Research and Reviews*, 2023. <https://doi.org/10.30574/wjarr.2023.20.2.1956>.
- [31] P. Hiskiawan, Y. M. Geasela, H. Heryanto, E. Stephanie, M. Ardianti, and F. A. Sukarno, "Trustworthy data science framework for non-invasive nutritional screening using computer vision," in *2025 International Conference on Informatics, Multimedia, Cyber and Information System (ICIMCIS)*, (Jakarta, Indonesia), pp. 1743–1748, 2025. <https://doi.org/10.1109/ICIMCIS68501.2025.11327268>.