Future Technology
Open Access Journal
ISSN 2832-0379

February 2027| Volume 06 | Issue 01 | Pages 77-92

Journal homepage: https://fupubco.com/futech

https://doi.org/10.55670/fpll.futech.6.1.7

Future Publishing LLC

Article

Predicting the direction of volatility in the Vietnamese stock market: A machine learning approach integrating macroeconomic uncertainty, financial stress, and petrodollar dynamics

Trinh Vo Thi Ngoc[1,2]*, Tuyen Le Nam[1,2], Nguyen Anh Phong[1,2], Tam Phan Huy[1,2], Quang Thuy Chu[1,2]

[1]University of Economics and Law, Ho Chi Minh City, Vietnam
[2]Vietnam National University Ho Chi Minh City, Vietnam

ARTICLE INFO

Article history:
Received 12 August 2026
Received in revised form
10 November 2026
Accepted 23 November 2026

Keywords:
Directional volatility forecasting,
Out-of-sample evaluation, GARCH-MIDAS
Economic policy uncertainty, Machine learning

*Corresponding author
Email address:
vtntrinh@uel.edu.vn

DOI: 10.55670/fpll.futech.6.1.7

ABSTRACT

This research forecasts the direction of next-month realized volatility in the Vietnamese stock market, a newly upgraded emerging market, and tests whether global policy uncertainty, geopolitical risk, a domestic financial stress index, and dollar-oil price changes add predictive content beyond market history. A walk-forward design on monthly data from 2015 to 2025 produces seventy-one out-of-sample forecasts for 2020 to 2025. Five machine learning classifiers and their average are compared with heterogeneous autoregressive (HAR), generalized autoregressive conditional heteroskedasticity mixed-data sampling (GARCH-MIDAS), and rank-based benchmarks. About two-thirds of months are classified correctly, and the best ensemble reaches an area under the curve of 0.726. A rule that only ranks current volatility in its own history reaches 0.713, so most of this ability reflects reversion from unusually high or low levels. No classifier beats the HAR benchmark. Adding uncertainty and stress raises (worsens) the Brier score by 0.002, and adding all four blocks by 0.007. When predictable reversion is removed from the target, no information set beats a constant forecast. The in-sample contribution of the blocks is significant under a Gaussian likelihood-ratio test but not under a test robust to fat tails. A timing rule based on market history raises the Sharpe ratio in point estimates, but only a specification identified after the fact is significant. In this market and sample, a lean model of market history is the more reliable basis for directional risk signals.

1. Introduction

Volatility-managed portfolios scale exposure inversely to expected risk [1], hedge ratios and margin requirements are set against expected volatility, and a rise in volatility can turn leverage into forced selling, so investors and supervisors often need to know whether volatility will rise or fall next month. Realized volatility (RV) clusters and reverts slowly to a long-run level, so its level is forecastable [2,3]. The sign of the next change also depends on the current level, the speed of reversion, and measurement noise, so information that helps forecast the level need not help forecast the direction. Volatility in emerging markets responds to external shocks and to domestic amplification. Economic policy uncertainty (EPU) and geopolitical risk (GPR) raise the option value of waiting and the risk premium [4–6]. Oil prices, United States inflation expectations, interest rates, and the dollar link commodity cycles to the global financial cycle that drives capital flows to emerging markets [7-10]; this research calls this market-price side of oil revenue recycling petrodollar dynamics. Domestic credit, funding and currency stress amplify shocks through balance sheets and liquidity spirals [11,12]. Vietnam, an MSCI frontier market that FTSE Russell reclassified as secondary emerging with effect from 21 September 2026 [13], is a demanding case: reliable monthly data cover about a decade, banking indicators are published with delays and across sources, and Lunar New Year closures produce short trading months. Mixed-frequency studies find that EPU, GPR, financial stress and oil shocks carry information about the long-run component of volatility [14-17], and flexible learners can match heterogeneous

autoregressive (HAR) benchmarks [18,19]. Three gaps remain. This literature mostly forecasts the level of volatility, while studies of the direction of volatility use United States data and volatility history alone [20,21]. Whether theoretically motivated macro-financial blocks add to volatility history in an emerging market with a short sample and regime shifts has not been tested. Predictors are usually added one at a time, so it is unclear whether the blocks add complementary information beyond market history. Variables that explain volatility in-sample often fail out-of-sample [22,23]. In Vietnam, machine learning has been applied to prices and returns [24,25], and the in-sample response to external uncertainty has been measured [26], but real-time directional forecasts of risk have not been tested.

The research problem is whether information that explains volatility in sample also improves real-time forecasts of the direction of risk in a market with a short history and shifting regimes. This research therefore evaluates whether the direction of monthly realized volatility of the Vietnamese stock market can be predicted out of sample, and whether macroeconomic uncertainty, financial stress, and petrodollar dynamics improve these forecasts beyond market history. Five research questions are addressed:

RQ1. Is the direction of next-month volatility predictable out of sample, and how much of this predictability is mechanical?

RQ2. Do machine learning classifiers outperform naive, rank-based, and econometric benchmarks?

RQ3. Do the macro-financial blocks improve the forecasts beyond market history, block by block and jointly?

RQ4. Which inputs do the fitted models rely on?

RQ5. Do the forecasts have economic value for volatility timing?

The research proposes no new algorithm and is designed as a test that can return a negative answer. In design, it evaluates all blocks within one walk-forward protocol against benchmarks that separate the mechanical predictability of the target, the information set, and the algorithm. Empirically, it finds that the macro-financial signal is weak and unstable both in and out of the sample. Methodologically, it shows that a rank-based benchmark accounts for most of the apparent directional skill and that a Gaussian likelihood-ratio test overstates the significance of GARCH-MIDAS drivers. The organization is as follows: (i) Section 1 introduces the topic, (ii) Section 2 reviews the literature, (iii) Section 3 describes data and models, (iv) Section 4 presents the results, and (v) Section 5 concludes.

2. Literature review

2.1 Background theories

Under the Heterogeneous Market Hypothesis (HMH), participants trade at different horizons and volatility cascades from long to short horizons [27]; the HAR model makes RV depend on its own short, medium, and long averages [3], and volatility dynamics are the main source of direction-of-change predictability [28]. Uncertainty raises the option value of waiting [4] and the risk premium [6]; geopolitical risk operates through trade [5]; and in GARCH-MIDAS models, such variables drive the slow, long-run level of volatility [29]. When net worth falls or funding tightens, external finance becomes dearer, and price shocks grow; the financial accelerator [11] and margin constraints force further sales [12]; a Financial Stress Index (FSI) summarizes this state [30]. In the global financial cycle, asset prices and capital flows share a factor driven by United States monetary policy and risk appetite [7, 31], and a stronger dollar tightens the balance sheets of dollar borrowers and global banks [8, 32]. Petrodollar dynamics refer to the recycling of oil revenues into dollar assets and the joint movement of oil prices, the dollar, and interest rates [9, 10]; this research uses market-price proxies, which measure global dollar and oil financial conditions rather than the flows themselves.

These theories describe the level of volatility, whereas the target is the sign of its next change, so Section 3.2 states two expected effects for each predictor. A high current RV lowers the probability of a further rise, partly by construction; given it, higher three- and six-month averages imply upward reversion, negative returns raise volatility through the leverage effect, and tighter dollar–oil financing conditions should raise the probability of a rise. For EPU, GPR, and the FSI, the direction is ambiguous: a rise ahead of turbulence makes an increase more likely, but a high level reached with or after a spike marks a state from which volatility tends to fall. RQ3 is tested through three hypotheses: adding the uncertainty (H1), stress (H2), or petrodollar (H3) block, alone or jointly, improves out-of-sample directional forecasts over market history. Under each null, the nested models of Section 3.2 have equal predictive accuracy. Three supplementary tests for RQ3 follow; they were formulated during revision, not registered in advance: (E1) increases in EPU, GPR and the FSI predict positive volatility shocks, that is, next-month RV above its HAR forecast; (E2) high EPU and FSI combined with high current RV predict a fall; and (E3) any predictive value of the blocks is concentrated in periods of financial stress.

2.2 Empirical studies

In GARCH-MIDAS studies of emerging economies, global EPU improves out-of-sample forecasts of the level of volatility, mainly in unstable periods [17]; GPR carries similar information [15, 16]; a global FSI outperforms the VIX, EPU and GPR [14], and oil shocks predict volatility in BRICS and international markets [33,34]. These studies forecast the level. Other evidence is more cautious: macro-financial variables explain United States volatility in sample but rarely beat autoregressive benchmarks out of sample [23], financial-market variables carry more content than macroeconomic aggregates [22], and during the pandemic the VIX helped most of nineteen markets whereas EPU helped few [35]. In emerging markets, the effects of uncertainty and stress are concentrated in extreme states [36], and in Vietnam trading partners' EPU matters more than GPR [26]. The studies disagree on whether in-sample gains carry over out of sample, whether uncertainty matters in general or only in extreme states, and whether machine learning gains at high frequency hold at low frequency. These disagreements track differences in target, frequency, and evaluation design, which this research holds fixed while it varies the information set.

Machine learning is competitive with the HAR family [18, 37] and gains from multi-channel predictors [19,38]; Vietnamese studies predict price trends and returns, not risk [24,25], and directional volatility studies use United States data [20,21] (Table 1). To the authors' knowledge, no study has tested, block by block and with in-sample and out-of-sample evidence side by side, whether macro-financial information improves monthly directional forecasts of realized volatility in a frontier or newly emerging market.

3. Methodology

3.1 Data

The research uses monthly data for Vietnam from January 2015 to December 2025 (132 observations). Daily VN-Index prices from Investing.com, collected from January 2013, are used to compute log returns and monthly RV over trading

sessions; prices from the second half of 2014 initialize the three- and six-month averages and momentum, and daily open, high, and low prices provide range-based RV for robustness. Lunar New Year months have as few as 14 sessions against an average of 20.8, so GARCH-MIDAS, which uses the 2,727 daily returns from February 2015, forecasts variance over the actual sessions, and a robustness design rescales RV to a common month of 21 sessions, $RV_t^* = RV_t(21/N_t)^{1/2}$, equivalent to multiplying realized variance by $21/N_t$; it changes the label in 16 of 131 months (4 out of sample). Because the label compares $RV_t/N_t^{1/2}$ across months and the constant shifts every log predictor equally, the labels, standardized predictors, and results are identical with the sample mean of 20.8. In each year, the Lunar New Year falls in a month with only 14 to 17 sessions; January averages 19.2 sessions and February 16.9, compared with 19.5 to 22.2 in the other months, while the annual means stay between 20.7 and 21.0.

Uncertainty is measured by the global EPU index of Davis [44], built with the method of Baker et al. [45], and by the global GPR index [5]. The FSI follows Park and Mercado [30] and averages expanding-window z-scores of the non-performing loan ratio, the lending–deposit spread, and USD/VND realized volatility. The loan ratio is annual to 2016, semiannual to mid-2023, and quarterly afterward; it enters from the quarter after its reference period and jumps in 2023Q3 (from 3.886 to 5.852 percent), where the source flags a break in definition, which coincides with the switch of the lending–deposit spread to International Monetary Fund reference rates; the spread is annual or half-yearly until mid-2023, so only the exchange-rate component changes every month. Petrodollar dynamics are proxied by monthly changes in the Brent oil price, the United States 10-year breakeven inflation rate, the 2-year Treasury yield, and the broad dollar index. Each row is for a month t uses values dated up to the end of the month t, and the target refers to $t+1$, leaving 131 usable months. For the extended sample from 2009, the FSI is spliced from year-end loan ratios and the previous year's average lending–deposit spread (over 2015–2025 the spliced index and the main index have a correlation of 0.914), and the three- and six-month averages and momentum are initialized with VN-Index prices from July to December 2008. Robust designs also use daily VNFIN, E1VFVN30, and VN30F1M prices from the VCI data service. The series are final revised data, and EPU, GPR, and some United States series are published after month-end, so the design is pseudo-real-time: no full real-time data vintages were used, and the lending–deposit spread enters in the months it refers to, with no publication lag; this favors the macro-financial blocks, and a design that lags them one more month removes it. Predictors are standardized with training-window statistics.

3.2 Empirical models

Let $r_{t,d}$ be the daily log return of the VN-Index on trading day d of month t. Following Andersen and Bollerslev [2], the monthly RV and the target, the direction of volatility (VOLUP), are:

$$RV_t = \left(\sum_{d=1}^{N_t} r_{t,d}^2\right)^{1/2} \quad (1)$$

$$VOLUP_t = 1(RV_{t+1} > RV_t) \quad (2)$$

where N_t is the number of trading days in month t. The sampling standard error of $lnRV_t$ is approximately $\sqrt{RQ_t/(2N_tRV_t^4)} = \sqrt{\sum_d r_{t,d}^4/(6RV_t^4)}$ with realized quarticity $RQ_t = (N_t/3)\sum_d r_{t,d}^4$ [46] (median 0.164), which separates clear from marginal changes. Three- and six-month averages capture the heterogeneous structure:

$$RV_t^{(k)} = \frac{1}{k}\sum_{j=0}^{k-1} RV_{t-j}, \quad k \in \{3,6\} \quad (3)$$

The predictor vector stacks four blocks, $X_t = [V_t, U_t, S_t, D_t]$: market history V_t (the logarithms of RV_t, $RV_t^{(3)}$ and $RV_t^{(6)}$, the monthly return and three- and six-month momentum), uncertainty U_t (the logarithms of EPU and GPR), stress S_t (the FSI) and petrodollar dynamics D_t (Table 2). Each classifier estimates:

$$p_t = Pr(VOLUP_t = 1 \mid X_t) = f(X_t) \quad (4)$$

and predicts a rise when $p_t > 0.5$, on nested information sets: market history alone, plus each block, plus uncertainty and stress, and all four blocks.

Table 1. Summary of selected empirical studies

Study	Market and data	Target and evaluation	Main finding
Yu et al. [17]	9 emerging markets; GARCH-MIDAS	Level, out of sample	Global EPU helps, mainly in unstable periods
Liang et al. [14]	21 international indices	Level, out of sample	Global FSI beats VIX, EPU and GPR
Paye [23]	United States, monthly	Level, in and out of sample	Macro variables explain in sample, rarely forecast
Das et al. [36]	24 emerging markets, monthly	Returns and volatility, in sample	Effects concentrated in extreme states
Christensen et al. [18]	Dow Jones stocks, high frequency	Level, out of sample	Machine learning competitive with HAR
Fernandes et al. [39]	United States, VIX, daily	Level of implied volatility	HAR with macro-financial variables fits well
Nyberg [40]	United States, monthly	Sign of excess returns	Some out-of-sample predictability
Phan et al. [41]	International, monthly	Returns with EPU	Predictive power varies across countries
Campisi et al. [42]	S&P 500	Direction of returns	Classifiers beat linear regression
Latif et al. [43]	S&P 500	Market direction and level	EPU, FSI, GPR, oil and technical inputs combined
Vrontos et al. [20]	United States, implied volatility	Direction of volatility	Predictable; machine learning better
Xie and Clements [21]	S&P 500 realized and implied	Direction of volatility	HAR-based classifiers improve forecasts
Cao and Vo [26]	Vietnam, 2000–2023	Volatility transmission, in sample	Partners' EPU matters more than GPR
Phuoc et al. [24]	Vietnam, VN-Index and VN30	Direction of price trend	High accuracy for price trends
Hoang et al. [25]	Vietnam, listed stocks	Returns	Return prediction, not risk

Note: Studies are grouped by target: the level of volatility, the direction or level of returns, the direction of volatility, and Vietnam.

Table 2. Variables measurement summary

Block	Variable (symbol)	Measurement	Level	Direction	ADF p	KPSS p	References
Target	Direction of volatility (VOLUP)	1 if realized volatility in month t+1 exceeds that in month t, 0 otherwise			—	—	[21,28]
Market history	Realized volatility, one month (lnRV)	Log of the square root of the sum of squared daily log returns in month t	+	− [b]	<0.001	>0.100	[2,3]
	Realized volatility, three months (lnRV3)	Log of average realized volatility over months t−2 to t	+	+	0.009	>0.100	[3,18]
	Realized volatility, six months (lnRV6)	Log of average realized volatility over months t−5 to t	+	+	0.055	>0.100	[3,21]
	Monthly return (RET)	ln P(t) − ln P(t−1) of the VN-Index	−	−	<0.001	>0.100	[47]
	Three-month momentum (MOM3)	ln P(t) − ln P(t−3)	−	− [c]	0.008	>0.100	[47,48]
	Six-month momentum (MOM6)	ln P(t) − ln P(t−6)	−	− [c]	0.079	>0.100	[47,48]
Macroeconomic uncertainty	Economic policy uncertainty (lnEPU)	Log of the global economic policy uncertainty index, current-price weights	+	± [a]	0.214	<0.010	[17,44]
	Geopolitical risk (lnGPR)	Log of the global geopolitical risk index	+	± [a]	0.007	0.017	[5,15]
Financial stress	Financial stress index (FSI)	Average of expanding-window z-scores of the non-performing loan ratio, lending–deposit spread and USD/VND volatility	+	± [a]	0.285	<0.010	[14,30]
Petrodollar dynamics (dollar–oil proxies)	Oil price change (ΔOIL)	Monthly log change of the Brent crude oil price	±	±	<0.001	>0.100	[9,33]
	Inflation expectations change (ΔBEI)	Monthly change in the United States ten-year breakeven inflation rate, percentage points	+	+	<0.001	>0.100	[9]
	Short-term yield change (ΔUS2Y)	Monthly change in the United States two-year Treasury yield, percentage points	+	+	<0.001	>0.100	[10,31]
	Dollar change (ΔUSD)	Monthly log change of the broad trade-weighted dollar index	+	+	<0.001	>0.100	[8,32]

Note: Level is the expected effect of a higher value on the level of realized volatility; Direction is the expected effect, conditional on the other market-history inputs, on the probability that volatility rises next month; petrodollar signs refer to monthly changes, not levels (once the three volatility terms are partialled out, the partial correlation with VOLUP is 0.210 for ΔUS2Y and at most 0.100 in absolute value for the other variables, compared with a band of ±0.171). Values outside this band appear only for ΔUS2Y with VOLUP (0.218) and with lnRV two and three months later (0.179 and 0.172), and for ΔUSD with VOLUP two months later (−0.188). ADF: Augmented Dickey-Fuller p-value (null of a unit root, lag length by the Akaike criterion); KPSS: Kwiatkowski-Phillips-Schmidt-Shin p-value (null of level stationarity; tabulated p-values bounded at 0.010 and 0.100). Sample: 132 months (131 for VOLUP). The signs predate estimation; E1–E3 were added in revision as supplementary tests. [a] Positive if the variable rises before volatility does; negative if a high level coincides with or follows a volatility spike (expectations E1 and E2). [b] Partly mechanical, because the target compares next month with current volatility. [c] The realized sign is opposite: VOLUP correlates positively with MOM3 (0.29) and MOM6 (0.20).

The first benchmark is the HAR model [3], estimated monthly by ordinary least squares on log RV; HAR-X adds macro-financial predictors Z_t:

$$lnRV_{t+1} = \beta_0 + \beta_1 lnRV_t + \beta_3 lnRV_t^{(3)} + \beta_6 lnRV_t^{(6)} + \gamma' Z_t + \varepsilon_{t+1} \quad (5)$$

and the probability of a rise is:

$$p_t^{HAR} = \Phi\left[\left(\widehat{lnRV}_{t+1} - lnRV_t\right)/\hat{\sigma}_\varepsilon\right] \quad (6)$$

where Φ is the standard normal distribution function. Because Φ is monotone with Φ(0) = 0.5, the predicted direction and AUC of HAR do not depend on normality, only on its Brier score. Section 4.2 compares these probabilities with a walk-forward recalibration.

The second is the GARCH-MIDAS model [29] on daily returns, with monthly drivers of the long-run component in GARCH-MIDAS-X. For day i of month m,

$$r_{i,m} = \mu + \sqrt{\tau_m g_{i,m}}\, z_{i,m}\,,$$

$$g_{i,m} = (1 - \alpha - \beta) + \alpha \frac{(r_{i-1,m} - \mu)^2}{\tau_m} + \beta g_{i-1,m} \quad (7)$$

$$ln\tau_m = c + \delta' Z_{m-1} \quad (8)$$

where $g_{i,m}$ is the short-run and τ_m the long-run component; each driver enters with one monthly lag because the sample is short. Estimation is by quasi-maximum likelihood (QML) with $\alpha, \beta > 0$, $\alpha + \beta < 1$, and robust standard errors [49], and joint tests of the drivers use the robust Wald statistic, because a Gaussian likelihood-ratio statistic is invalid under

fat tails. Next-month variance $H_{t+1} = \tau_{t+1} \sum_{h=0}^{N_{t+1}-1} [1 + (\alpha + \beta)^h (g_1 - 1)]$ is corrected by b_t, the average of $ln(RV_m^2 / \hat{H}_m)$ in the estimation sample, and a rise is predicted when $H_{t+1} e^{b_t} > RV_t^2$.

Simple rules separate the mechanical predictability of the target from the value of information and algorithm: the majority class, persistence of the last change, and the rank rule, which treats RV as independent over time, $\hat{p}_t = \#\{s \leq t : RV_s > RV_t\}/t$, and predicts a rise when current volatility is below its historical median; logits on lnRV and on the HAR terms and a HAR regression on six inputs complete the set. Every model is also compared with the rank rule by the difference in AUC and the Brier skill score $1 - BS/BS^{rank}$. A second target removes the mechanical component: with the HAR regression (5) without Z_t,

$$SHOCK_t = 1\left(lnRV_{t+1} > \hat{E}_t[lnRV_{t+1}]\right) \quad (9)$$

where the HAR forecast of each month is estimated on the months before it (at least 24). Training and test labels thus share one pseudo-out-of-sample definition; a version with in-sample residual labels is reported in Section 4.3. A model that predicts SHOCK must know something the HAR model does not. The five classifiers are logistic regression with a mild ridge penalty, the elastic net [50], random forest [51], XGBoost and LightGBM [52]; the ensemble averages their probabilities, and the linear classifiers use the same six inputs as the trees. For the logistic models:

$$p_t = \Lambda(b_0 + b' X_t),$$

$$\hat{b} = argmin_b \left\{ -l(b) + \lambda \left[\rho \parallel b \parallel_1 + \frac{1-\rho}{2} \parallel b \parallel_2^2 \right] \right\} \quad (10)$$

where Λ is the logistic function and l the log-likelihood. Baseline hyperparameters are conservative values carried over from the project's earlier classifiers for the direction of monthly returns on the same sample, so they were not tuned on this target but are not fully independent of the evaluation period: inverse penalties of 1 and 0.5 (equal lasso and ridge weights), 500 trees of depth 4 with at least five observations per leaf, and 300 boosted trees of depth 3 with a learning rate of 0.03, subsampling of 0.8 and an L2 penalty of 2. These values were set in September 2026 for the return classifiers, which used the same 2015–2025 sample and forecast origins; no alternatives were searched, and the values were not changed after the volatility forecasts were produced. Four checks bracket them (Section 4.3): more regularized (inverse penalties of 0.1; depth 3 for the forest; depth 2 with 200 boosted trees) and more flexible sets (penalties of 10 and 2; depth 6; depth 4 with 500 trees and a learning rate of 0.05), a set chosen per classifier by forward-chained log loss on 2009–2014 with market history only and no evaluation data, and per-origin re-tuning by log loss over three forward-chained folds of the training window on a three-point grid per classifier: inverse penalties {0.1, 1, 10} for the logit and {0.1, 0.5, 2} for the elastic net, forest depths {3, 4, 6}, and boosted depth and learning rate {(2, 0.03), (3, 0.03), (4, 0.05)}; the carried-over values remain the baseline so that the forecasts match the pre-specified design, and the 2009–2014 choice leads to the same conclusions (Section 4.3). Tree probabilities are calibrated with Platt scaling [53] by three-fold cross-fitting inside each training window; calibration is kept because the Brier score and the timing rule use the probabilities, and a design without it is reported. At each origin t from January 2020 to November 2025, models are trained on all earlier origins (at least sixty) and forecast $t + 1$, giving seventy-one out-of-sample forecasts that include the pandemic crash, the 2022 sell-off and the 2025 tariff shock; GARCH-MIDAS is re-estimated at every origin, and a rolling sixty-month window is a robustness check. Directional accuracy (DA), the macro-averaged F1 score, the area under the receiver operating characteristic curve (AUC), and the Brier score (BS) are reported:

$$DA = \frac{1}{T} \sum_t 1\left(\widehat{VOLUP}_t = VOLUP_t\right)$$

$$BS = \frac{1}{T} \sum_t (p_t - VOLUP_t)^2 \quad (11)$$

The Pesaran and Timmermann [54] (PT) test checks directional independence, and the Diebold and Mariano [55] (DM) test, with the correction of Harvey et al. [56] compares Brier losses, $d_t = L_t^A - L_t^B$:

$$DM = \frac{\bar{d}}{\sqrt{\hat{V}(\bar{d})}} \sqrt{\frac{T-1}{T}} \quad (12)$$

against the one-sided alternative that model B is more accurate. Because DM has no standard limit for nested models on expanding windows, the conditional Giacomini-White [57] (GW) test with instruments $(1, d_{t-1})$ is applied to both window designs. AUC differences use the DeLong test [58], 95 percent intervals come from a stationary bootstrap [59], and the minimum detectable effect of each DM test at 5 percent size and 80 percent power is $(z_{0.95} + z_{0.80})\, \hat{\sigma}_d / \sqrt{T}$. The 90 percent model confidence set (MCS) [60] collects the models whose Brier loss cannot be distinguished from the best; the fluctuation test of Giacomini and Rossi [61] over 24 forecasts checks for time variation. The Clark-West test [62] was derived for nested linear forecasts under squared error, so it is only a supplementary check for calibrated tree probabilities; the GW test, valid with a finite estimation window, is the formal nested test. With 71 forecasts, the binomial standard error of a DA near 0.65 is about 0.057. The primary comparisons are the market-history ensemble against HAR and the augmented ensembles against it; a non-rejection means no detectable improvement, not a zero effect. Economic value is measured with an idealized timing rule on the non-traded VN-Index, holding the index in month $t + 1$ when a fall in volatility is predicted and bank deposits otherwise:

$$R_{t+1}^s = w_t R_{t+1} + (1 - w_t) R_t^f - c_b (w_t - w_{t-1})^+ - c_s (w_{t-1} - w_t)^+,$$

$$w_t = 1(p_t \leq 0.5) \quad (13)$$

where R_t^f is the average deposit rate for terms up to twelve months, from State Bank of Vietnam and Banking Review publications to mid-2023 and the International Monetary Fund reference deposit rate afterward, divided by twelve; Sharpe ratios and the Jobson-Korkie and Ledoit-Wolf tests use returns in excess of it. The baseline sets $c_b = c_s = 0.2$ percent, a round trip of 0.4 percent matching brokerage fees plus the 0.1 percent tax on sales; checks use 0.15 and 0.25 percent, next-day execution, because month-t inputs are fully known only after the close, and a continuous weight $w_t = 1 - p_t$; the same signals are also applied to the E1VFVN30 exchange-traded fund and VN30F1M futures, delaying trades when the ±7 percent price limit binds and allowing for T+2 settlement. Strategies are compared with buy-and-hold and with a static portfolio of the same average exposure. Equal Sharpe ratios are tested with the robust Ledoit and Wolf test [63] and by permuting positions 5,000 times, singly and in six-month blocks, and the superior predictive ability (SPA) test

[64] controls for the search over ten strategies. Predictions of the tree ensembles are decomposed out of sample with Shapley Additive Explanations (SHAP) [65], $f(x) = \phi_0 + \sum_j \phi_j$, by variable and block, and each block's contribution to the ensemble's Brier score is decomposed by the Shapley value over the eight nested sets.

3.3 Data construction and code

All transformations use data dated up to the forecast origin. Each FSI component $x_{j,t}$ is standardized as $z_{j,t} = (x_{j,t} - m_{j,t})/s_{j,t}$, with expanding means and standard deviations from January 2015 to month t, and $FSI_t = (1/3)\Sigma_j z_{j,t}$. The loan ratio for reference period q enters from the first month of the quarter after q; annual values for 2015–2016 are carried through the following year; the International Monetary Fund series, used from 2017Q2, is not back-spliced across the 2023Q3 break; and the first twelve months are standardized with the 2015 mean and standard deviation.

Platt scaling maps a tree score f to $p = 1/[1 + \exp(Af + B)]$, with A and B fitted by maximum likelihood on out-of-fold scores from three stratified, unshuffled folds of each training window, so calibration uses no data after the forecast origin. The code, the FSI components, and all derived series are in the replication package described under Data availability.

4. Results and discussion

4.1 Descriptive analysis

RV averages 4.735 percent per month and rose in 52.7 percent of the 131 months and 47.9 percent of the out-of-sample months. The Augmented Dickey-Fuller (ADF) test [66] rejects a unit root at the 1 percent level for most variables and at the 10 percent level for lnRV6 (p = 0.055) and six-month momentum (p = 0.079); the KPSS test [67] does not reject stationarity for these, so the marginal ADF results reflect low power. Both tests point to drift in lnEPU and the FSI (ADF p = 0.214 and 0.285, KPSS p < 0.010; Table 2), and for lnGPR the KPSS test rejects as well (p = 0.017), so their changes are also tested. VOLUP correlates with current volatility at −0.38. Inside the blocks, the highest correlations are 0.84 for the three- and six-month volatility averages, 0.71 for the momentum terms, and 0.51 for the oil and breakeven changes. After 2019, RV rises from 4.042 to 5.312 percent, log EPU from 5.120 to 5.531, and the FSI from −0.733 to 0.230, with its standard deviation rising from 0.469 to 1.063 (both differences significant at 1 percent), partly because of the 2023 definition change.

4.2 Forecasting results

Table 3 reports the out-of-sample forecasts. The naive rules fail, while the rank rule, the logits, HAR, GARCH-MIDAS, and all market-history classifiers pass the PT test at 5 percent, and HAR, GARCH-MIDAS, and the market-history ensemble remain significant after a Holm correction. Much of this comes from the construction of the target. The rank rule reaches an AUC of 0.713; only the market-history ensemble (by 0.012), the logit on lnRV, and the elastic net exceed it, none significantly (DeLong p = 0.773 for the ensemble), and if RV were independent over time, a HAR model would reach an AUC of 0.830. Market-history models mainly improve probabilities, with Brier skill scores of 0.095 for HAR, 0.108 for the ensemble and 0.140 for a logit on lnRV (DM p = 0.021 against the rank rule). For RQ1, the direction is predictable about two months in three, largely because unusual volatility reverts towards normal. For the market-history ensemble, the 95 percent stationary-bootstrap intervals (Table 3) are 0.563 to 0.746 for the DA and 0.631 to 0.818 for the AUC (HAR: 0.549 to 0.732 and 0.584 to 0.806), and the balanced accuracy, which weights rises and falls equally, is 0.670 against 0.500 for the majority rule. Errors concentrate where the label is uncertain: in the 45 months whose change in lnRV exceeds the combined quarticity-based standard error, the ensemble is correct in 0.800 of months (AUC 0.873; HAR 0.756, GARCH-MIDAS 0.778), but in the 26 marginal months all models lie between 0.423 and 0.654, so a threshold target raises accuracy without changing the ranking of models or information sets. Without the eleven months that have the largest absolute changes, the ensemble AUC is still 0.715 (0.695 and 0.670 with added blocks), and when one year of forecasts is dropped at a time, it stays between 0.717 and 0.745, so predictability does not depend on a few extreme months.

For RQ2, the market-history ensemble has the highest AUC (0.726), the logit on lnRV has the lowest Brier score (0.212), and GARCH-MIDAS and XGBoost have the highest accuracy (0.676). No classifier beats HAR significantly (smallest DM p among eighteen specifications 0.253), linear classifiers match the trees, and the 90 percent MCS retains 24 of 25 probability models, so the data cannot rank most specifications. No advantage of the algorithm over a parsimonious model of recent volatility is detected. Recalibrating the HAR probabilities walk-forward lowers its Brier score from 0.223 to 0.216 (isotonic) or 0.220 (logistic), level with the ensemble (DM p = 0.416; Figure 1), so the normal approximation does not handicap HAR. Table 4 answers RQ3. Adding uncertainty and stress raises the accuracy of three of five classifiers but lowers the AUC of all five and improves the Brier score of none; no block alone raises any classifier's AUC, and with all four blocks every model is worse. The ensemble's accuracy falls by 0.099 with uncertainty because 27 percent of its probabilities lie within 0.05 of 0.5, so a mean change of 0.033 flips seven predictions. None of thirty nested DM tests and no DeLong test favors an augmented model, and the GW tests and rolling window agree (Table 4, Panel B). Large gains can be excluded, because the minimum detectable Brier improvement is only 0.015 to 0.022, or about 7 to 10 percent of the score.

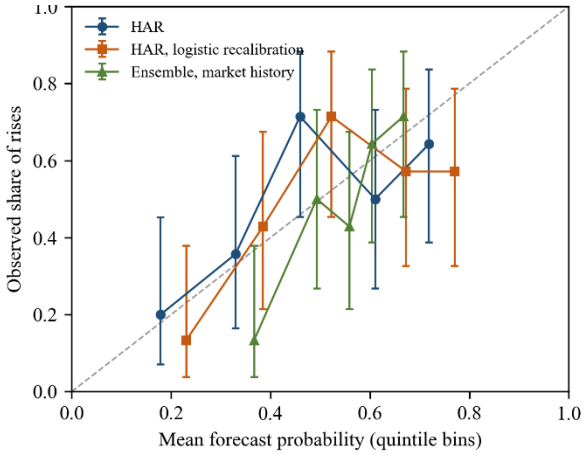


Figure 1. Reliability of the HAR and market-history ensemble probabilities

Note: The seventy-one forecasts are sorted into quintile bins of the forecast probability, and the observed share of rises is shown with 95 percent Wilson intervals. Brier scores are 0.223 for HAR, 0.220 after walk-forward logistic recalibration (0.216 with isotonic recalibration), and 0.220 for the market-history ensemble.

Table 3. Out-of-sample directional forecasts of next-month volatility

Model	DA	DA 95% CI	F1	AUC	AUC 95% CI	ΔAUC vs rank	Brier	BSS vs rank	PT p-value
Panel A. Benchmarks									
Majority class	0.479	–	0.324	–	–	–	–	–	–
Persistence (last change)	0.437	–	0.436	0.436	–	−0.277	–	–	0.863
Rank rule (i.i.d. benchmark)	0.620	0.521–0.704	0.609	0.713	0.598–0.835	0.000	0.247	0.000	0.022
Logit, lnRV only	0.634	–	0.634	0.719	–	0.006	0.212	0.140	0.011
HAR	0.648	0.549–0.732	0.648	0.693	0.584–0.806	−0.020	0.223	0.095	0.006
HAR-X (uncertainty and stress)	0.563	–	0.558	0.630	–	−0.083	0.247	0.000	0.152
HAR-X (all blocks)	0.535	–	0.516	0.629	–	−0.085	0.254	−0.031	0.307
GARCH-MIDAS	0.676	0.592–0.761	0.675	0.696	0.596–0.807	−0.017	–	–	0.001
GARCH-MIDAS-X (uncertainty and stress)	0.634	–	0.633	0.669	–	−0.045	–	–	0.012
GARCH-MIDAS-X (all blocks)	0.634	–	0.632	0.625	–	−0.089	–	–	0.012
Panel B. Machine learning: market history									
Logistic regression	0.620	0.521–0.718	0.619	0.708	0.613–0.810	−0.005	0.221	0.104	0.017
Elastic net	0.648	0.535–0.746	0.647	0.718	0.615–0.825	0.004	0.213	0.135	0.005
Random forest	0.592	0.507–0.690	0.563	0.665	0.544–0.775	−0.049	0.233	0.055	0.018
XGBoost	0.676	0.577–0.775	0.663	0.657	0.546–0.749	−0.057	0.240	0.028	<0.001
LightGBM	0.592	0.493–0.690	0.556	0.608	0.477–0.717	−0.105	0.245	0.008	0.013
Ensemble (average)	0.662	0.563–0.746	0.654	0.726	0.631–0.818	0.012	0.220	0.108	0.001

Note: Seventy-one monthly forecasts from origins January 2020 to November 2025. The 95 percent intervals are percentiles of 5,000 stationary-bootstrap resamples (mean block length four) and are computed only for the benchmarks used in the main comparisons (–: not computed). Two benchmarks are not shown: the logit on the HAR terms (DA 0.648, AUC 0.700, Brier 0.221) and the HAR regression on six market-history inputs (DA 0.662, AUC 0.693, Brier 0.227). The classifiers with added blocks are compared in Table 4. ΔAUC vs rank is the difference from the rank rule's AUC, BSS vs rank is the Brier skill score 1 − Brier/0.247, and PT p-value is the one-sided Pesaran-Timmermann p-value. The GARCH-MIDAS models produce point forecasts, so the Brier score is not reported, and the AUC is computed from the forecast log-ratio of next-month to current variance.

Smaller gains cannot be ruled out, and a two- to three-point gain in DA, roughly half a binomial standard error, would be detected in only 16 to 21 percent of samples. With lnEPU, lnGPR, and the FSI in monthly changes, which removes their drift, the augmented sets are still worse (Section 4.3), and in the clear-change months the ordering is unchanged (0.873, 0.819, and 0.720).

The supplementary tests detect no support for the expectations (Section 4.3). Changes in EPU, GPR, and the FSI do not predict the HAR residual (E1; joint Wald p = 0.545). No information set forecasts the volatility shock of equation (9) better than a constant (smallest DM p = 0.700; Section 4.3). With all blocks, the AUC falls to 0.322, which is significantly below 0.5 (p = 0.006), as is the set with monthly changes (0.362, p = 0.041). In-sample residual labels give the same ordering (0.482, 0.420 and 0.394), so the result is not an artifact of the labels; an AUC below 0.5 means that relations learned before 2020 reversed afterward (Section 4.4). Return direction shows the same pattern (ensemble AUC 0.454, and 0.363 and 0.311 with added blocks): with 46 rises and 25 falls, the null standard error is about 0.072, so 0.311 is below 0.5 at 1 percent (p ≈ 0.009).

An audit found no coding error: the labels refer to month t + 1, the probabilities refer to a rise, and the low AUC comes from the two linear classifiers (0.339 and 0.313), which are not calibrated, while the three trees stay close to 0.5 (0.480 to 0.531). The reversal therefore reflects unstable learned relations. Once the predictable reversion is removed, no information in these data is found to forecast the direction of volatility. The interactions of lnRV with EPU and the FSI are insignificant (E2; p = 0.117 and 0.472), and the augmented ensembles do slightly better in months of high FSI, but not detectably (E3; conditional GW p = 0.109 and 0.179).

The in-sample evidence is weak as well (Table 5). In the HAR regressions, only current volatility is significant, and the added blocks are not (Wald p = 0.740 and 0.355). In GARCH-MIDAS, adding EPU, GPR and the FSI gives a Gaussian likelihood-ratio statistic of 25.077 (p < 0.001) and lowers the Bayesian information criterion from −17162.6 to −17163.9, but the robust Wald test does not reject (p = 0.208); with the petrodollar block the criterion worsens to −17143.6 and the robust tests are insignificant (p = 0.135; petrodollar alone p = 0.260). Only EPU is positive at 10 percent (δ = 0.313, robust standard error 0.180).

Table 4. Contribution of each information block relative to market history, precision and power of the primary comparisons, and SHAP shares within blocks

Information set	ΔDA	ΔAUC	ΔBrier	DM p	DeLong p	Higher DA	Higher AUC	Lower Brier	Min. DM p
Panel A. Augmented minus market-history ensemble									
+ Uncertainty	−0.099	−0.033	0.005	0.821	0.261	2 of 5	0 of 5	1 of 5	0.486
+ Stress	−0.042	−0.001	−0.001	0.428	0.954	3 of 5	0 of 5	3 of 5	0.426
+ Petrodollar	−0.070	−0.025	0.003	0.693	0.582	2 of 5	0 of 5	1 of 5	0.497
+ Uncertainty and stress	−0.042	−0.023	0.002	0.633	0.428	3 of 5	0 of 5	0 of 5	0.509
+ All blocks	−0.070	−0.057	0.007	0.764	0.271	0 of 5	0 of 5	0 of 5	0.695
Panel B. Precision and power of the primary comparisons									
Comparison	ΔBrier, 95% CI		MDE	DM power	ΔAUC, 95% CI		DeLong p	GW p, expand.	GW p, rolling
MH ensemble vs HAR	[−0.026, 0.021]		0.042	–	[−0.045, 0.102]		0.491	0.345	0.964
MH ensemble vs rank rule	[−0.076, 0.017]		0.067	–	[−0.076, 0.089]		0.773	–	–
+ Uncertainty and stress	[−0.011, 0.014]		0.015	0.950	[−0.092, 0.041]		0.428	0.963	0.198
+ All blocks	[−0.010, 0.024]		0.022	0.729	[−0.167, 0.030]		0.271	0.627	0.375
Panel C. SHAP shares of the inputs in each nested specification (rank among all inputs in parentheses)									
Input	+ U	+ S	+ D	+ U&S	All	+ U&S, changes		All, changes	
lnRV (reference)	0.310 (1)	0.340 (1)	0.293 (1)	0.295 (1)	0.235 (1)	0.276 (1)		0.227 (1)	
lnEPU or ΔlnEPU	0.145 (2)	–	–	0.130 (2)	0.119 (2)	0.155 (2)		0.122 (2)	
lnGPR or ΔlnGPR	0.061 (7)	–	–	0.057 (8)	0.028 (13)	0.074 (7)		0.056 (7)	
FSI or ΔFSI	–	0.120 (4)	–	0.101 (5)	0.078 (5)	0.081 (6)		0.066 (6)	
ΔOIL	–	–	0.050 (9)	–	0.035 (11)	–		0.042 (11)	
ΔBEI	–	–	0.079 (5)	–	0.062 (7)	–		0.049 (9)	
ΔUS2Y	–	–	0.079 (6)	–	0.064 (6)	–		0.077 (5)	
ΔUSD	–	–	0.080 (4)	–	0.054 (9)	–		0.043 (10)	

Note: MH: market history. Panel A: differences between the augmented and market-history ensembles (a positive ΔBrier means worse probabilities); DM p is the one-sided Diebold-Mariano and DeLong p the two-sided DeLong p-value for the ensemble. Counts refer to the five classifiers; Min. DM p is the smallest DM p-value over them and the ensemble. Panel B: 95 percent intervals for the differences (second model minus first; the point estimates are in Table 3 and Panel A), from 5,000 stationary-bootstrap resamples with a mean block length of four. MDE is the smallest Brier improvement that the one-sided DM test detects with 80 percent power at 5 percent. Power is the rejection rate at 5 percent after a Brier reduction of 0.02 is added to the demeaned loss differences; the corresponding GW power is 0.887 and 0.587, and the DM power for a DA gain of 0.03 is 0.210. The GW tests use the instruments (1, d(t−1)). Over the 30 comparisons of Panel A, the Clark-West p-values range from 0.278 to 0.772. Panel C: mean absolute out-of-sample SHAP value of each input divided by the sum over all inputs, averaged over the uncalibrated random forest, XGBoost and LightGBM refitted at each origin; U is uncertainty, S stress and D petrodollar dynamics; in the changes columns lnEPU, lnGPR and the FSI enter as monthly changes; –: not in the set.

With fat-tailed daily returns, a Gaussian test can thus overstate the significance of GARCH-MIDAS drivers. Estimates converge at every origin, $\alpha + \beta$ lies between 0.953 and 0.989, and squared standardized residuals show no dependence (Ljung-Box p-values of 0.233 to 0.673 at ten lags), but the standardized residuals themselves show weak autocorrelation (p = 0.044 to 0.062). On samples from July 2015 (January 2016 for the twelve-month polynomial), two monthly lags per driver or a restricted beta-lag polynomial over six or twelve months give a worse BIC than one lag on the same sample (Table 5, Panel C), the second lags are jointly insignificant (robust Wald p = 0.662), and the beta weights of EPU and GPR fall almost entirely on the first lag. In this shorter sample, the EPU coefficient is significant at about 5 percent in every lag structure (t = 1.957 to 2.177), but the joint robust tests of the drivers do not reject (p = 0.084 to 0.278).

Figure 2 reports the SHAP attributions (RQ4). In the market-history trees, current RV receives 37.6 percent of the attribution and three-month momentum 16.9 percent; higher volatility lowers, and higher momentum raises the predicted probability of a rise, the latter against the expected sign. In the all-block trees, current RV still leads (23.5 percent).

The macro-financial inputs receive 44.0 percent when the variable shares in panel (b) are added, and 51.2 percent when the SHAP values of each block are summed before taking absolute values (grouped SHAP); yet their Shapley contributions to the ensemble's Brier score are +0.004 for uncertainty and petrodollar dynamics (worse) and −0.002 for stress, which is not significant in any test. Higher EPU and FSI values lower the probability of a rise, contrary to their expected effect on volatility levels (Figure 2c–d; Section 4.4). In monthly changes, EPU and GPR attributions rise with their values in all three tree ensembles (correlations of 0.58 to 0.71 and 0.51 to 0.54; Figure 2e–f), whereas the FSI change stays negative (−0.56 to −0.53) because its monthly component, USD/VND volatility, moves with stock volatility. Within blocks (Table 4, Panel C), lnEPU takes more than twice the share of lnGPR whenever both enter (0.145 against 0.061 with uncertainty alone), and with every block lnGPR has the smallest share of the thirteen inputs (0.028); the dollar, breakeven and two-year yield changes split the petrodollar attribution almost equally (0.079 to 0.080 when that block is added on its own), with oil lowest (0.050). Only the uncertainty block thus reduces to one variable.

Table 5. In-sample estimates of the econometric benchmarks

Parameter	Market history only	+ Uncertainty and stress	+ All blocks
Panel A. HAR regressions for next-month log realized volatility			
lnRV	0.197*** (4.273)	0.200*** (4.396)	0.203*** (4.791)
Adjusted R^2	0.228	0.217	0.221
Wald test, added blocks (p-value)	–	1.252 (0.740)	7.754 (0.355)
Panel B. GARCH-MIDAS models for daily returns (long-run component drivers)			
μ (percent per day)	0.071*** (3.460)	0.068*** (3.457)	0.069*** (3.674)
α	0.132*** (5.123)	0.132*** (6.041)	0.134*** (6.427)
β	0.831*** (20.406)	0.840*** (28.888)	0.838*** (30.914)
c (long-run constant)	−8.778*** (−37.244)	−8.688*** (−25.375)	−8.713*** (−25.889)
δ lnEPU		0.313* (1.740)	0.314* (1.777)
δ lnGPR		−0.063 (−0.339)	−0.085 (−0.434)
δ FSI		−0.065 (−0.367)	−0.032 (−0.179)
Log-likelihood	8597.1	8609.6	8615.3
BIC	−17162.6	−17163.9	−17143.6
Gaussian likelihood-ratio test (p-value)	–	25.077 (<0.001)	36.438 (<0.001)
Robust Wald test, added drivers (p-value)	–	4.544 (0.208)	11.085 (0.135)
Panel C. GARCH-MIDAS + uncertainty and stress, alternative lag structures			
Lag structure	Two lags	Beta, K = 6	Beta, K = 12
δ lnEPU, first lag	0.357** (2.177)	0.327* (1.957)	0.313** (1.969)
δ lnGPR, first lag	−0.124 (−0.662)	−0.131 (−0.662)	−0.157 (−0.798)
δ FSI, first lag	−0.138 (−0.648)	0.071 (0.466)	0.118 (0.775)
BIC minus one-lag BIC	17.1	23.2	22.1
Robust Wald, second lags (p)	1.587 (0.662)	–	–

Note: Panel A: ordinary least squares on 131 months with Newey-West [68] standard errors (three lags) and standardized predictors; the other slope coefficients, and the petrodollar δ coefficients of Panel B are not significant at 10 percent (the largest absolute t-statistic is 1.389). Panel B: quasi-maximum likelihood on 2,727 daily returns with Bollerslev-Wooldridge robust t-statistics; δ coefficients are effects of a one-standard-deviation change in the lagged monthly driver on the log long-run variance. The robust Wald test uses the sandwich covariance of the δ coefficients. t-statistics in parentheses; ***, ** and * denote significance at the 1, 5 and 10 percent levels. Panel C uses daily returns from July 2015 (January 2016 for K = 12); the beta weights are proportional to (1 − k/(K + 1))^(w − 1), k = 1, …, K, with w in [1, 60]. The estimated w of 59.688 (K = 6) and 60.000 (K = 12) for lnEPU and 60.000 for lnGPR places nearly all of the weight on the first lag. The BIC difference is measured against the one-lag model on the same sample, so a positive value means that one lag is preferred.

Table 6 and Figure 3 evaluate the timing rule (RQ5). The pre-specified market-history ensemble raises the Sharpe ratio from 0.385 to 0.557 and halves the maximum drawdown while holding the index in 32.4 percent of months, also with next-day execution (0.701). The gain is not significant (Ledoit-Wolf p = 0.332, block permutation p = 0.162) and depends on March 2020, without which the ensemble's Sharpe ratio is 0.564 against 0.649 for buy-and-hold; a static portfolio with the same exposure has a smaller drawdown (−0.086), and the alpha against it is 4.7 percent a year (t = 0.949). XGBoost with market history passes all tests (Ledoit-Wolf p = 0.006) but was selected after the fact; the SPA test rejects over ten strategies (p = 0.036) but not over the five pre-specified ones (p = 0.610). Asymmetric costs change Sharpe ratios by less than 0.01. On tradable instruments, the ensemble does not beat buy-and-hold (Sharpe ratios 0.449 against 0.604 for the E1VFVN30 fund and 0.626 against 0.700 for VN30 futures), while XGBoost keeps 1.330 and 1.397. For RQ5, no pre-specified strategy shows detectable economic value.

4.3 Robustness tests

Table 7 brings together all re-estimated designs, the volatility-shock target and the alternative proxies. Predictability survives every design that keeps the target (market-history AUCs 0.612 to 0.740). It is weakest for range-based RV [70, 71], whose labels agree with the baseline in 87 percent of forecast months for the Parkinson and 82 percent for the Garman–Klass estimator, and which the ±7 percent price limits also truncate. The 21-session rescaling and uncertainty and stress in changes leave the ordering intact, and the extended sample from 2009 (143 forecasts) confirms predictability (market-history AUC 0.665, DA 0.671) without a detectable gain from the blocks (smallest DM p = 0.240); against HAR its Clark-West p-value is 0.051 and its DM p-value 0.493, and from 2015, when the deposit series starts, the timing rule earns a Sharpe ratio of 0.214 against 0.366 for buy-and-hold (Ledoit-Wolf p = 0.772). The hyperparameters chosen on 2009–2014 yield the highest market-history AUC and the smallest gaps, so the baseline values do not favor market history; per-origin re-tuning selects depth 2 for XGBoost and LightGBM in 86 and 93 percent of origins, depth 3 for the forest in 58 percent of origins, and the strongest logistic penalty at every origin. Freezing the loan ratio at its 2023Q2 value or dropping the 11 January and February targets leaves the ordering unchanged. No re-estimated design shows a detectable improvement from the original blocks (smallest p = 0.107 over six models).

Alternative proxies change little (Panel C). The VIX [35] does not help; EPU weighted by Vietnam's 2019 trade with its five main partners beats global EPU (p = 0.087) but not market history, and Treasury holdings of Saudi Arabia, the United Arab Emirates and Kuwait (TIC data, lagged two months), a direct measure of petrodollar recycling, add nothing. A market-priced FSI built from USD/VND volatility and the volatility and beta of the VNFIN financial-sector index (available from mid-2017) is the only proxy with a nominally significant gain: it raises the AUC of all five classifiers and lowers the ensemble's Brier score by 0.005 (DM p = 0.049, GW p = 0.038, DeLong p = 0.288). The gain does not survive a Holm adjustment over the seven proxy designs (p = 0.341) and disappears when uncertainty is added, but it suggests that timely market prices, not slow banking ratios, are where stress information lies.

4.4 Discussion

A weak and unstable signal, in sample and out of sample: The evidence fits the small predictable component, estimation noise and unstable relations stressed by Paye [23], and weak proxies. The long-run component driven by EPU, GPR and the FSI explains 4.4 percent of the variance of monthly changes in log realized variance and 12.3 percent of its level (9.8 and 15.9 percent with all four blocks), and the

predictors left their training range: in 21.1 percent of forecast months at least one of lnEPU, lnGPR and the FSI lies outside its expanding-window range, and in 28.2 percent at least one macro-financial input does; the FSI alone is outside its range in 9.9 percent of months, all between December 2021 and November 2023, and its loan-ratio component in 5.6 percent, compared with 2.8 percent for both the spread and USD/VND volatility. In-sample relations shift (Figure 4b): the EPU coefficient reaches 0.363 only at the April 2025 tariff shock, the FSI coefficient is positive only from September 2023 to February 2025, and the GPR coefficient ranges from −0.395 to −0.060.

Out of sample (Figure 4a), the all-block ensemble is correct in only 0.458 of months in 2020–2021, against 0.667 for market history, but is the most accurate model during the domestic bond-market and banking stress of 2022–2023 (0.708), consistent with the state dependence reported by Das et al. [36] (Table 8). Within periods, the conditional GW test rejects equal accuracy only in 2020–2021 (p = 0.021), where market history is better; over the whole period these differences are not detectably significant (fluctuation statistic within ±2.190 against a 10 percent critical value of 2.756).

Table 6. Economic value of volatility-timing strategies

Strategy	CAGR	Volatility	Sharpe ratio	Max. DD	Exposure	LW p-value	Random p-value	Block p-value	Sharpe, next day	Pos. changes	Break-even cost
Buy-and-hold	0.115	0.232	0.385	−0.328	1.000	–	–	–	0.387	1	–
HAR switch	0.080	0.200	0.252	−0.255	0.507	0.742	0.440	0.454	0.308	29	–
GARCH-MIDAS switch	0.066	0.161	0.181	−0.273	0.535	0.686	0.540	0.550	0.143	27	–
Ensemble, market history	0.120	0.136	0.557	−0.167	0.324	0.332	0.138	0.162	0.701	21	0.87%
Ensemble, + uncertainty and stress	0.072	0.193	0.217	−0.249	0.451	0.782	0.463	0.488	0.353	25	–
Ensemble, all blocks	−0.007	0.162	−0.247	−0.348	0.394	0.983	0.903	0.921	−0.135	25	–
XGBoost, market history	0.178	0.091	1.335	−0.041	0.282	0.006	<0.001	<0.001	1.324	17	3.18%

Note: Idealized strategies hold the VN-Index when a fall in volatility is predicted and bank deposits otherwise, with a cost of 0.2 percent per one-way change of position, February 2020 to December 2025. CAGR is the compound annual return, volatility is the annualized volatility, Max. DD is the maximum drawdown, and Exposure is the share of months in the index. The Sharpe ratio is in excess of the deposit rate. LW p-value is from the one-sided robust Ledoit-Wolf test against buy-and-hold; Random and Block p-values are the shares of 5,000 permutations of monthly positions, singly or in six-month blocks, with a Sharpe ratio at least as high. Sharpe, next-day trades at the first close of the following month. Pos. changes counts switches over the 71 months; the break-even cost is the cost per change at which the Sharpe ratio falls to that of buy-and-hold at the baseline cost (–: below buy-and-hold even without costs). Average holding period in the index: 2.400 months for HAR, 2.714 for GARCH-MIDAS, 2.091, 2.462, and 2.154 for the three ensembles, and 2.222 for XGBoost. The Jobson-Korkie [69] test gives p-values of 0.322 for the market-history ensemble and 0.019 for XGBoost.

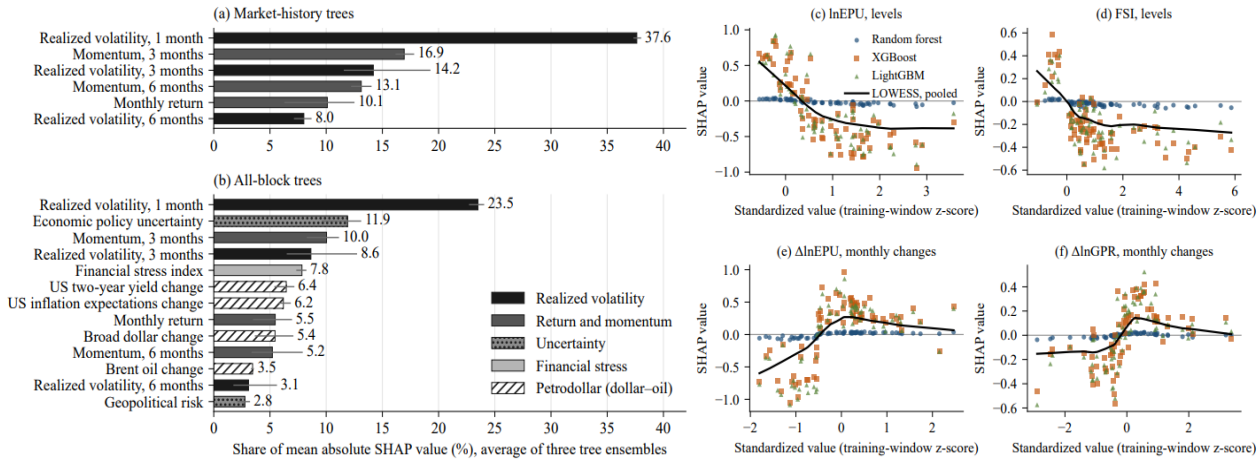


Figure 2. Out-of-sample Shapley attributions of the tree ensembles: (a) market history, (b) all four blocks, (c)–(d) SHAP dependence of the levels of EPU and the FSI, and (e)–(f) of the monthly changes in EPU and GPR

Note: Bars show the average share across random forest, XGBoost, and LightGBM, and lines show the range; panel (a) uses market history only, and panel (b) uses all four blocks. Panels (c)–(f) plot the out-of-sample SHAP values of the three tree models against the training-window z-score of the input, with all four blocks in levels (c, d) or with uncertainty and stress in monthly changes (e, f); the line is a LOWESS fit pooled over the three models.

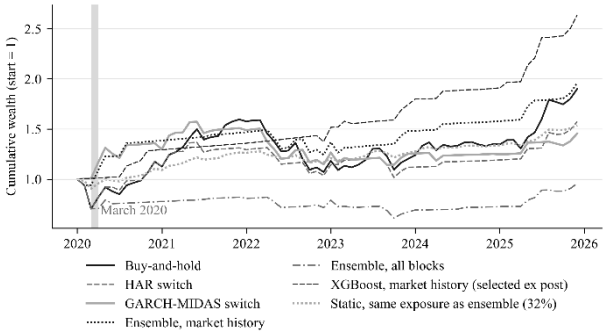


Figure 3. Cumulative wealth of volatility-timing strategies, February 2020 to December 2025
Note: The shaded band marks March 2020; the static portfolio is rebalanced monthly to its fixed exposure and shown after costs.

The blocks also fail for the level of volatility under squared-error and QLIKE losses [72]: the lowest QLIKE loss (0.617) belongs to GARCH-MIDAS without drivers, and adding the blocks increases the losses of both econometric models. Level versus direction, the negative attributions of EPU and the FSI: The SHAP signs describe how the models use the inputs, not causal effects on the level of volatility [4, 6]. In the top tercile of EPU, RV averages 5.471 percent and rises next month in only 0.386 of cases, against 0.636 in the bottom tercile (Table 9); for the FSI, the shares are 0.386 and 0.614. Conditional on the three volatility terms, the top-minus-bottom gap in the share of rises narrows from −0.250 to −0.192 for EPU (p = 0.060) and from −0.227 to −0.194 for the FSI (p = 0.055), so volatility history absorbs part of the pattern but not all of it. In a logit for the direction, lnEPU has a coefficient of −0.271 (p = 0.033) alone but −0.232 (p = 0.204) with the volatility terms, so its association weakens once volatility history is controlled for.

Table 7. Re-estimated designs, volatility-shock target and alternative proxies

Design	Reference AUC	Augmented AUC (1)	Augmented AUC (2)	Best ΔBrier	Min. DM p
Panel A. Re-estimated designs (reference: market history)					
Baseline (expanding window, 2015–2025)	0.726	0.703	0.669	0.002	0.509
Macro-financial inputs lagged one more month	0.740	0.688	0.672	0.004	0.235
Range-based RV, Parkinson	0.688	0.677	0.604	0.002	0.107
RV rescaled to 21 sessions	0.714	0.698	0.676	0.003	0.110
Uncertainty and stress in monthly changes	0.726	0.670	0.674	0.006	0.453
FSI from spread and USD/VND volatility only	0.726	0.703	0.669	0.003	0.182
FSI with loan ratio frozen at 2023Q2	0.726	0.701	0.669	0.003	0.522
Excluding January and February targets	0.733	0.712	0.691	0.004	0.561
Extended sample 2009–2025 (143 forecasts)	0.665	0.679	0.677	−0.000	0.240
Rolling sixty-month window	0.667	0.625	0.618	0.008	0.455
Range-based RV, Garman–Klass	0.612	0.580	0.556	0.006	0.537
No probability calibration	0.715	0.680	0.680	0.009	0.622
Hyperparameters chosen on 2009–2014	0.732	0.725	0.696	0.000	0.464
Hyperparameters re-tuned at each origin	0.725	0.715	0.700	0.001	0.475
Hyperparameters more regularized	0.723	0.719	0.699	0.000	0.454
Hyperparameters more flexible	0.715	0.690	0.644	0.004	0.453
High-volatility target	0.695	0.647	0.614	0.001	0.372
Panel B. Volatility-shock target, pseudo-out-of-sample labels					
Market history; augmented (1) uncertainty and stress, (2) all blocks	0.459	0.438	0.322	0.005	0.278
In-sample residual labels	0.482	0.420	0.394	0.008	0.267
Panel C. Alternative proxies (reference: market history)					
+ VIX	0.726	0.718	–	0.002	0.367
Trade-weighted EPU replaces global EPU, with stress	0.726	0.712	–	0.000	0.457
+ Treasury holdings of oil exporters, with petrodollar block	0.726	0.707	–	0.005	0.666
+ Market-priced FSI	0.726	0.743	–	−0.005	0.005

Note: Ensemble AUCs. In Panels A and B, augmented (1) adds uncertainty and stress and (2) all four blocks; in Panel C, (1) is the set with the alternative proxy. Best ΔBrier is the smallest difference in Brier score between an augmented and the reference ensemble (positive means worse); Min. DM p is the smallest one-sided Diebold-Mariano p-value over five classifiers and the ensemble. In Panel B a constant forecast has a Brier score of 0.254 and the reference ensemble 0.264. For the extended sample, Min. DM p is the smallest over all eighteen nested comparisons. The design without January and February targets keeps the baseline forecasts and drops 11 of them. In the five hyperparameter designs (baseline, chosen on 2009–2014, re-tuned, more regularized, more flexible), the market-history DA is 0.662, 0.662, 0.634, 0.662 and 0.634, the market-history Brier score 0.220, 0.219, 0.223, 0.222 and 0.223, and the smallest Clark-West p-value 0.299, 0.266, 0.184, 0.266 and 0.257. Section 3.2 lists the re-tuning grids. The rolling window and the high-volatility target use the same seventy-one forecast months.

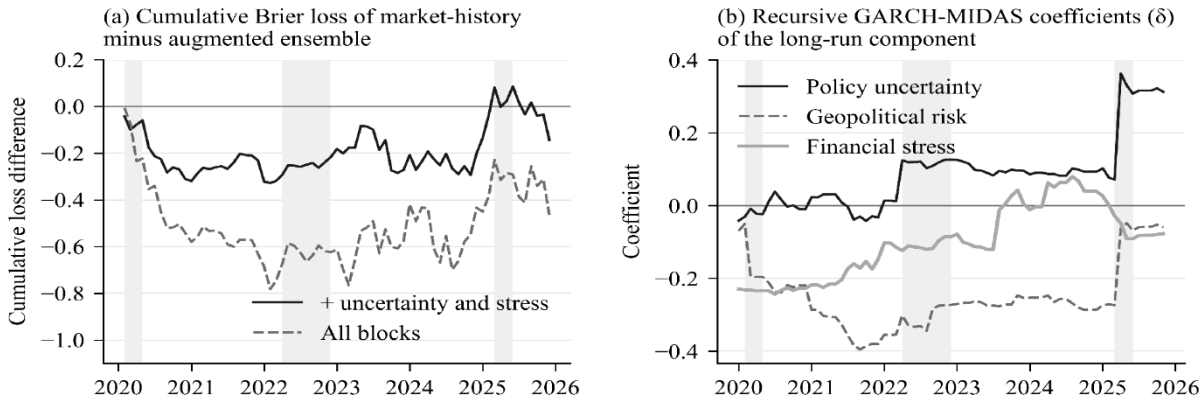


Figure 4. Out-of-sample losses and recursive in-sample coefficients of the macro-financial blocks, 2020 to 2025
Note: Panel (a): cumulative Brier loss of the market history minus the augmented ensembles (a rise means the augmented model is better). Panel (b): recursive GARCH-MIDAS coefficients with uncertainty and stress. Shaded areas mark February–April 2020, April–November 2022, and March–May 2025.

Table 8. Sub-period performance of the market-history and all-block ensembles

Origins	n	DA MH	DA all	AUC MH	AUC all	Brier MH	Brier all	GW p
2020–2021	24	0.667	0.458	0.706	0.476	0.228	0.256	0.021
2022–2023	24	0.667	0.708	0.779	0.829	0.205	0.194	0.542
2024–2025	23	0.652	0.609	0.674	0.644	0.228	0.230	0.324

Note: Periods are defined by forecast origin. GW p is the conditional Giacomini-White p-value for equal Brier loss of the two ensembles within the period, with instruments (1, dt−1); with 23 or 24 forecasts, the test has little power. Clark-West p-values for the all-block against the market-history ensemble are 0.996, 0.132, and 0.355.

Table 9. Next-month change in realized volatility and share of rises by terciles of lnEPU and the FSI

Tercile	RV in month t (%)	Mean ΔRV, t to t+1 (pp)	Share of rises	Share of rises, conditional
Panel A. lnEPU				
Low	4.408	0.118	0.636	0.099
Middle	4.276	0.040	0.558	−0.007
High	5.471	−0.107	0.386	−0.093
High minus low	–	–	−0.250 (0.019)	−0.192 (0.060)
Panel B. FSI				
Low	4.534	0.351	0.614	0.077
Middle	4.610	−0.060	0.581	0.040
High	5.018	−0.241	0.386	−0.116
High minus low	–	–	−0.227 (0.033)	−0.194 (0.055)

Note: 131 labeled months, split into terciles of the contemporaneous value (44, 43 and 44 months). The conditional share is the mean residual of VOLUP from an ordinary least squares regression on lnRV, lnRV3 and lnRV6, which is zero on average. The p-values in parentheses come from Welch t-tests of the top against the bottom tercile.

The FSI's cross-correlation with RV is near zero when the index leads (average −0.028) and exceeds the approximate band of ±0.174 only when volatility leads by 14 to 23 months, consistent with an index that follows turbulence as loan losses are recognized [73]; five of the eight months with the highest readings, all between July 2023 and April 2024, were followed by a fall in volatility.

5. Conclusions

For this market, sample and set of proxies, three conclusions follow. First, the direction of volatility is predictable out of sample, mostly through reversion of unusually high or low volatility, and flexible learners do not beat HAR or a one-variable logit, complementing United States evidence; with the same inputs, the direction of returns is not predictable (AUC 0.454). Second, no consistent improvement from the macro-financial blocks is detected in or out of sample, although the tests can only exclude large gains; the mixed-frequency studies reporting gains forecast the level at higher frequency, and, as in United States predictive regressions, a weak and unstable signal and imperfect proxies fit the findings, with timely market-priced stress indicators the most promising. Because the composite FSI changes every month only through USD/VND volatility, the banking-stress channel is tested only weakly; the monthly market-priced FSI is the sharper test, and its nominal gain does not survive the multiplicity adjustment. Third, timing

rules raise the Sharpe ratio in point estimates, but no pre-specified rule shows detectable value, on the index or on tradable instruments. Two methodological lessons apply beyond Vietnam: directional volatility forecasts should be judged against a rank-based benchmark, and GARCH-MIDAS drivers should be tested with statistics robust to fat tails. Whether United States directional accuracy survives analogous macro-financial blocks is left for future work. For market participants and regulators, a monthly probability of rising volatility can enter exposure limits, hedge ratios or a monitoring dashboard as one input, not a rule for margins, and a richer model should first beat HAR and a rank rule out of sample and net of costs. The research does not forecast the size of changes or tail risk, so it gives no basis for thresholds, and timely market-priced indicators would suit a dashboard better than the loan ratio. The research has limitations. With seventy-one forecasts the minimum detectable improvement in Brier score is 0.015 to 0.022, and the design is pseudo-real-time; longer histories or panels of markets would help. Interbank and bond spreads, credit default swaps, margin debt, and foreign flows could not be assembled for 2015–2025. RV from about 21 daily closes is noisy and truncated by price limits in crises; GARCH-MIDAS uses one monthly lag per driver, and the cash return is a deposit rate rather than a one-month interbank rate.

Acknowledgements
This research is funded by Vietnam National University HoChiMinh City (VNU-HCM) under a project within the framework of the Program titled "Strengthening the capacity for education and basic scientific research integrated with strategic technologies at VNU-HCM, aiming to achieve advanced standards comparable to regional and global levels during the 2025-2030 period, with a vision toward 2045".

Ethical issue
The authors are aware of and comply with best practices in publication ethics, specifically with regard to authorship (avoidance of guest authorship), dual submission, manipulation of figures, competing interests, and compliance with policies on research ethics. The authors adhere to publication requirements that the submitted work is original and has not been published elsewhere.

Data availability statement
The manuscript contains all the data. However, additional data will be provided by the corresponding author upon reasonable request. All data are from the public sources described in Section 3.1. A replication package with the code, processed data, data dates, seeds, and the output files that reproduce every reported number is deposited in the Open Science Framework (OSF) at https://osf.io/k23yv/overview?view_only=bf10ddd261dd404c9ce0addb64090f77.

Conflict of interest
The authors declare no potential conflict of interest.

References

[1] A. Moreira and T. Muir, "Volatility-managed portfolios," The Journal of Finance, vol. 72, no. 4, pp. 1611–1644, 2017. DOI: https://doi.org/10.1111/jofi.12513

[2] T. G. Andersen and T. Bollerslev, "Answering the skeptics: Yes, standard volatility models do provide accurate forecasts," International Economic Review, vol. 39, no. 4, pp. 885–905, 1998. DOI: https://doi.org/10.2307/2527343

[3] F. Corsi, "A simple approximate long-memory model of realized volatility," Journal of Financial Econometrics, vol. 7, no. 2, pp. 174–196, 2009. DOI: https://doi.org/10.1093/jjfinec/nbp001

[4] N. Bloom, "The impact of uncertainty shocks," Econometrica, vol. 77, no. 3, pp. 623–685, 2009. DOI: https://doi.org/10.3982/ecta6248

[5] D. Caldara and M. Iacoviello, "Measuring geopolitical risk," American Economic Review, vol. 112, no. 4, pp. 1194–1225, 2022. DOI: https://doi.org/10.1257/aer.20191823

[6] Ľ. Pástor and P. Veronesi, "Uncertainty about government policy and stock prices," The Journal of Finance, vol. 67, no. 4, pp. 1219–1264, 2012. DOI: https://doi.org/10.1111/j.1540-6261.2012.01746.x

[7] H. Rey, "Dilemma not trilemma: The global financial cycle and monetary policy independence," in Proceedings of the Economic Policy Symposium, Jackson Hole, Kansas City, MO, USA: Federal Reserve Bank of Kansas City, 2013, pp. 285–333. [Online]. Available: https://fedinprint.org/item/fedkpr/21390

[8] V. Bruno and H. S. Shin, "Capital flows and the risk-taking channel of monetary policy," Journal of Monetary Economics, vol. 71, pp. 119–132, 2015. DOI: https://doi.org/10.1016/j.jmoneco.2014.11.011

[9] L. Kilian and X. Zhou, "Oil prices, exchange rates and interest rates," Journal of International Money and Finance, vol. 126, Art. no. 102679, 2022. DOI: https://doi.org/10.1016/j.jimonfin.2022.102679

[10] R. Koepke, "Fed policy expectations and portfolio flows to emerging markets," Journal of International Financial Markets, Institutions and Money, vol. 55, pp. 170–194, 2018. DOI: https://doi.org/10.1016/j.intfin.2018.03.003

[11] B. S. Bernanke, M. Gertler and S. Gilchrist, "The financial accelerator in a quantitative business cycle framework," in Handbook of Macroeconomics, vol. 1C, J. B. Taylor and M. Woodford, Eds. Amsterdam, Netherlands: Elsevier, 1999, pp. 1341–1393. DOI: https://doi.org/10.1016/S1574-0048(99)10034-X

[12] M. K. Brunnermeier and L. H. Pedersen, "Market liquidity and funding liquidity," The Review of Financial Studies, vol. 22, no. 6, pp. 2201–2238, 2009. DOI: https://doi.org/10.1093/rfs/hhn098

[13] FTSE Russell, "FTSE Russell announces results of March 2026 semi-annual country classification review – equities and fixed income," London Stock Exchange Group, 7 April 2026. [Online]. Available: https://www.lseg.com/en/media-centre/press-releases/ftse-russell/2026/ftse-russell-announces-results-march-2026-semi-annual-country-classification-review-equities-fixed-income [Accessed 04 Nov 2026].

[14] C. Liang, Q. Luo, Y. Li and L. D. T. Huynh, "Global financial stress index and long-term volatility forecast for international stock markets," Journal of International Financial Markets, Institutions and Money, vol. 88, Art. no. 101825, 2023. DOI: https://doi.org/10.1016/j.intfin.2023.101825

[15] A. A. Salisu, A. E. Ogbonna, L. Lasisi and A. Olaniran, "Geopolitical risk and stock market volatility in emerging markets: A GARCH–MIDAS approach," The North American Journal of Economics and Finance, vol. 62, Art. no. 101755, 2022. DOI: https://doi.org/10.1016/j.najef.2022.101755

[16] M. Segnon, R. Gupta and B. Wilfling, "Forecasting stock market volatility with regime-switching GARCH-MIDAS: The role of geopolitical risks," International Journal of Forecasting, vol. 40, no. 1, pp. 29–43, 2024. DOI: https://doi.org/10.1016/j.ijforecast.2022.11.007

[17] X. Yu, Y. Huang and K. Xiao, "Global economic policy uncertainty and stock volatility: Evidence from emerging economies," Journal of Applied Economics, vol. 24, no. 1, pp. 416–440, 2021. DOI: https://doi.org/10.1080/15140326.2021.1953913

[18] K. Christensen, M. Siggaard and B. Veliyev, "A machine learning approach to volatility forecasting," Journal of Financial Econometrics, vol. 21, no. 5, pp. 1680–1727, 2023. DOI: https://doi.org/10.1093/jjfinec/nbac020

[19] Z. Niu, C. Wang and H. Zhang, "Forecasting stock market volatility with various geopolitical risks categories: New evidence from machine learning models," International Review of Financial Analysis, vol. 89, Art. no. 102738, 2023. DOI: https://doi.org/10.1016/j.irfa.2023.102738

[20] S. D. Vrontos, J. Galakis and I. D. Vrontos, "Implied volatility directional forecasting: A machine learning approach," Quantitative Finance, vol. 21, no. 10, pp. 1687–1706, 2021. DOI: https://doi.org/10.1080/14697688.2021.1905869

[21] X. Xie and A. E. Clements, "Predicting directional volatility: HAR model with machine learning integration," Applied Economics Letters, vol. 33, no. 7, pp. 961–969, 2026. DOI: https://doi.org/10.1080/13504851.2024.2401512

[22] C. Christiansen, M. Schmeling and A. Schrimpf, "A comprehensive look at financial volatility prediction by economic variables," Journal of Applied Econometrics, vol. 27, no. 6, pp. 956–977, 2012. DOI: https://doi.org/10.1002/jae.2298

[23] B. S. Paye, "'Déjà vol': Predictive regressions for aggregate stock market volatility using macroeconomic variables," Journal of Financial Economics, vol. 106, no. 3, pp. 527–546, 2012. DOI: https://doi.org/10.1016/j.jfineco.2012.06.005

[24] T. Phuoc, P. T. K. Anh, P. H. Tam and C. V. Nguyen, "Applying machine learning algorithms to predict the stock price trend in the stock market – The case of Vietnam," Humanities and Social Sciences Communications, vol. 11, Art. no. 393, 2024. DOI: https://doi.org/10.1057/s41599-024-02807-x

[25] T. N. Hoang, N. Do-Thi and T. Nguyen-Trang, "Predicting stock returns using machine learning combined with data envelopment analysis and automatic feature engineering: A case study on the Vietnamese stock market," PLOS ONE, vol. 20, no. 9, Art. no. e0332154, 2025. DOI: https://doi.org/10.1371/journal.pone.0332154

[26] P. T.-H. Cao and D. H. Vo, "Market responses to geopolitical risk and economic policy uncertainty: Evidence from Vietnam," Heliyon, vol. 11, no. 4, Art. no. e42703, 2025. DOI: https://doi.org/10.1016/j.heliyon.2025.e42703

[27] U. A. Müller, M. M. Dacorogna, R. D. Davé, R. B. Olsen, O. V. Pictet and J. E. von Weizsäcker, "Volatilities of different time resolutions: Analyzing the dynamics of market components," Journal of Empirical Finance, vol. 4, no. 2–3, pp. 213–239, 1997. DOI: https://doi.org/10.1016/s0927-5398(97)00007-8

[28] P. F. Christoffersen and F. X. Diebold, "Financial asset returns, direction-of-change forecasting, and volatility dynamics," Management Science, vol. 52, no. 8, pp. 1273–1287, 2006. DOI: https://doi.org/10.1287/mnsc.1060.0520

[29] R. F. Engle, E. Ghysels and B. Sohn, "Stock market volatility and macroeconomic fundamentals," The Review of Economics and Statistics, vol. 95, no. 3, pp. 776–797, 2013. DOI: https://doi.org/10.1162/REST_a_00300

[30] C.-Y. Park and R. V. Mercado Jr., "Determinants of financial stress in emerging market economies," Journal of Banking & Finance, vol. 45, pp. 199–224, 2014. DOI: https://doi.org/10.1016/j.jbankfin.2013.09.018

[31] S. Miranda-Agrippino and H. Rey, "U.S. monetary policy and the global financial cycle," The Review of Economic Studies, vol. 87, no. 6, pp. 2754–2776, 2020. DOI: https://doi.org/10.1093/restud/rdaa019

[32] S. Avdjiev, V. Bruno, C. Koch and H. S. Shin, "The dollar exchange rate as a global risk factor: Evidence from investment," IMF Economic Review, vol. 67, no. 1, pp. 151–173, 2019. DOI: https://doi.org/10.1057/s41308-019-00074-4

[33] A. A. Salisu and R. Gupta, "Oil shocks and stock market volatility of the BRICS: A GARCH-MIDAS approach," Global Finance Journal, vol. 48, Art. no. 100546, 2021. DOI: https://doi.org/10.1016/j.gfj.2020.100546

[34] A. A. Salisu, A. E. Ogbonna, R. Gupta and E. Bouri, "Energy-related uncertainty and international stock market volatility," The Quarterly Review of Economics and Finance, vol. 95, pp. 280–293, 2024. DOI: https://doi.org/10.1016/j.qref.2024.04.005

[35] J. Wang, X. Lu, F. He and F. Ma, "Which popular predictor is more useful to forecast international stock markets during the coronavirus pandemic: VIX vs EPU?" International Review of Financial Analysis, vol. 72, Art. no. 101596, 2020. DOI: https://doi.org/10.1016/j.irfa.2020.101596

[36] D. Das, M. Kannadhasan and M. Bhattacharyya, "Do the emerging stock markets react to international economic policy uncertainty, geopolitical risk and financial stress alike?" The North American Journal of Economics and Finance, vol. 48, pp. 1–19, 2019. DOI: https://doi.org/10.1016/j.najef.2019.01.008

[37] C. Zhang, Y. Zhang, M. Cucuringu and Z. Qian, "Volatility forecasting with machine learning and intraday commonality," Journal of Financial Econometrics, vol. 22, no. 2, pp. 492–530, 2024. DOI: https://doi.org/10.1093/jjfinec/nbad005

[38] J. Luo, T. Klein, T. Walther and Q. Ji, "Forecasting realized volatility of crude oil futures prices based on machine learning," Journal of Forecasting, vol. 43, no. 5, pp. 1422–1446, 2024. DOI: https://doi.org/10.1002/for.3077

[39] M. Fernandes, M. C. Medeiros and M. Scharth, "Modeling and predicting the CBOE market volatility index," Journal of Banking & Finance, vol. 40, pp. 1–10, 2014. DOI: https://doi.org/10.1016/j.jbankfin.2013.11.004

[40] H. Nyberg, "Forecasting the direction of the US stock market with dynamic binary probit models," International Journal of Forecasting, vol. 27, no. 2, pp. 561–578, 2011. DOI: https://doi.org/10.1016/j.ijforecast.2010.02.008

[41] D. H. B. Phan, S. S. Sharma and V. T. Tran, "Can economic policy uncertainty predict stock returns? Global evidence," Journal of International Financial Markets, Institutions and Money, vol. 55, pp. 134–150, 2018. DOI: https://doi.org/10.1016/j.intfin.2018.04.004

[42] G. Campisi, S. Muzzioli and B. De Baets, "A comparison of machine learning methods for predicting the direction of the US stock market on the basis of volatility indices," International Journal of Forecasting, vol. 40, no. 3, pp. 869–880, 2024. DOI: https://doi.org/10.1016/j.ijforecast.2023.07.002

[43] S. Latif, F. Aslam, P. Ferreira and S. Iqbal, "Integrating macroeconomic and technical indicators into forecasting the stock market: A data-driven approach," Economies, vol. 13, no. 1, Art. no. 6, 2025. DOI: https://doi.org/10.3390/economies13010006

[44] S. J. Davis, "An index of global economic policy uncertainty," NBER Working Paper No. 22740, National Bureau of Economic Research, Cambridge, MA, USA, 2016. DOI: https://doi.org/10.3386/w22740

[45] S. R. Baker, N. Bloom and S. J. Davis, "Measuring economic policy uncertainty," The Quarterly Journal of Economics, vol. 131, no. 4, pp. 1593–1636, 2016. DOI: https://doi.org/10.1093/qje/qjw024

[46] O. E. Barndorff-Nielsen and N. Shephard, "Econometric analysis of realized volatility and its use in estimating stochastic volatility models," Journal of the Royal Statistical Society: Series B (Statistical Methodology), vol. 64, no. 2, pp. 253–280, 2002. DOI: https://doi.org/10.1111/1467-9868.00336

[47] J. Y. Campbell and L. Hentschel, "No news is good news: An asymmetric model of changing volatility in stock returns," Journal of Financial Economics, vol. 31, no. 3, pp. 281–318, 1992. DOI: https://doi.org/10.1016/0304-405x(92)90037-x

[48] N. Jegadeesh and S. Titman, "Returns to buying winners and selling losers: Implications for stock market efficiency," The Journal of Finance, vol. 48, no. 1, pp. 65–91, 1993. DOI: https://doi.org/10.1111/j.1540-6261.1993.tb04702.x

[49] T. Bollerslev and J. M. Wooldridge, "Quasi-maximum likelihood estimation and inference in dynamic models with time-varying covariances," Econometric Reviews, vol. 11, no. 2, pp. 143–172, 1992. DOI: https://doi.org/10.1080/07474939208800229

[50] H. Zou and T. Hastie, "Regularization and variable selection via the elastic net," Journal of the Royal Statistical Society: Series B (Statistical Methodology), vol. 67, no. 2, pp. 301–320, 2005. DOI: https://doi.org/10.1111/j.1467-9868.2005.00503.x

[51] L. Breiman, "Random forests," Machine Learning, vol. 45, no. 1, pp. 5–32, 2001. DOI: https://doi.org/10.1023/a:1010933404324

[52] T. Chen and C. Guestrin, "XGBoost: A scalable tree boosting system," in Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining, San Francisco, CA, USA, 2016, pp. 785–794. DOI: https://doi.org/10.1145/2939672.2939785

[53] J. C. Platt, "Probabilistic outputs for support vector machines and comparisons to regularized likelihood methods," in Advances in Large Margin Classifiers, A. J. Smola, P. Bartlett, B. Schölkopf and D. Schuurmans, Eds. Cambridge, MA, USA: MIT Press, 1999, pp. 61–74.

[54] M. H. Pesaran and A. Timmermann, "A simple nonparametric test of predictive performance," Journal of Business & Economic Statistics, vol. 10, no. 4, pp. 461–465, 1992. DOI: https://doi.org/10.1080/07350015.1992.10509922

[55] F. X. Diebold and R. S. Mariano, "Comparing predictive accuracy," Journal of Business & Economic Statistics, vol. 13, no. 3, pp. 253–263, 1995. DOI: https://doi.org/10.1080/07350015.1995.10524599

[56] D. Harvey, S. Leybourne and P. Newbold, "Testing the equality of prediction mean squared errors," International Journal of Forecasting, vol. 13, no. 2, pp. 281–291, 1997. DOI: https://doi.org/10.1016/s0169-2070(96)00719-4

[57] R. Giacomini and H. White, "Tests of conditional predictive ability," Econometrica, vol. 74, no. 6, pp. 1545–1578, 2006. DOI: https://doi.org/10.1111/j.1468-0262.2006.00718.x

[58] E. R. DeLong, D. M. DeLong and D. L. Clarke-Pearson, "Comparing the areas under two or more correlated receiver operating characteristic curves: A nonparametric approach," Biometrics, vol. 44, no. 3, pp. 837–845, 1988. DOI: https://doi.org/10.2307/2531595

[59] D. N. Politis and J. P. Romano, "The stationary bootstrap," Journal of the American Statistical Association, vol. 89, no. 428, pp. 1303–1313, 1994. DOI: https://doi.org/10.1080/01621459.1994.10476870

[60] P. R. Hansen, A. Lunde and J. M. Nason, "The model confidence set," Econometrica, vol. 79, no. 2, pp. 453–497, 2011. DOI: https://doi.org/10.3982/ECTA5771

[61] R. Giacomini and B. Rossi, "Forecast comparisons in unstable environments," Journal of Applied Econometrics, vol. 25, no. 4, pp. 595–620, 2010. DOI: https://doi.org/10.1002/jae.1177

[62] T. E. Clark and K. D. West, "Approximately normal tests for equal predictive accuracy in nested models," Journal of Econometrics, vol. 138, no. 1, pp. 291–311,

2007. DOI: https://doi.org/10.1016/j.jeconom.2006.05.023
[63] O. Ledoit and M. Wolf, "Robust performance hypothesis testing with the Sharpe ratio," Journal of Empirical Finance, vol. 15, no. 5, pp. 850–859, 2008. DOI: https://doi.org/10.1016/j.jempfin.2008.03.002
[64] P. R. Hansen, "A test for superior predictive ability," Journal of Business & Economic Statistics, vol. 23, no. 4, pp. 365–380, 2005. DOI: https://doi.org/10.1198/073500105000000063
[65] S. M. Lundberg and S.-I. Lee, "A unified approach to interpreting model predictions," in Advances in Neural Information Processing Systems 30, Long Beach, CA, USA, 2017, pp. 4765–4774. [Online]. Available: https://proceedings.neurips.cc/paper/2017/hash/8a20a8621978632d76c43dfd28b67767-Abstract.html
[66] D. A. Dickey and W. A. Fuller, "Distribution of the estimators for autoregressive time series with a unit root," Journal of the American Statistical Association, vol. 74, no. 366a, pp. 427–431, 1979. DOI: https://doi.org/10.1080/01621459.1979.10482531
[67] D. Kwiatkowski, P. C. B. Phillips, P. Schmidt and Y. Shin, "Testing the null hypothesis of stationarity against the alternative of a unit root: How sure are we that economic time series have a unit root?" Journal of Econometrics, vol. 54, no. 1–3, pp. 159–178, 1992. DOI: https://doi.org/10.1016/0304-4076(92)90104-Y
[68] W. K. Newey and K. D. West, "A simple, positive semi-definite, heteroskedasticity and autocorrelation consistent covariance matrix," Econometrica, vol. 55, no. 3, pp. 703–708, 1987. DOI: https://doi.org/10.2307/1913610
[69] J. D. Jobson and B. M. Korkie, "Performance hypothesis testing with the Sharpe and Treynor measures," The Journal of Finance, vol. 36, no. 4, pp. 889–908, 1981. DOI: https://doi.org/10.1111/j.1540-6261.1981.tb04891.x
[70] M. Parkinson, "The extreme value method for estimating the variance of the rate of return," The Journal of Business, vol. 53, no. 1, pp. 61–65, 1980. DOI: https://doi.org/10.1086/296071
[71] M. B. Garman and M. J. Klass, "On the estimation of security price volatilities from historical data," The Journal of Business, vol. 53, no. 1, pp. 67–78, 1980. DOI: https://doi.org/10.1086/296072
[72] A. J. Patton, "Volatility forecast comparison using imperfect volatility proxies," Journal of Econometrics, vol. 160, no. 1, pp. 246–256, 2011. DOI: https://doi.org/10.1016/j.jeconom.2010.03.034
[73] R. Beck, P. Jakubík and A. Piloiu, "Key determinants of non-performing loans: New evidence from a global sample," Open Economies Review, vol. 26, no. 3, pp. 525–550, 2015. DOI: https://doi.org/10.1007/s11079-015-9358-8



This article is an open-access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (https://creativecommons.org/licenses/by/4.0/).