ISA Journal of Business, Economics and Management (ISAJBEM)

Homepage: https://isapublisher.com/isajbem/

Volume 3, Issue 4, Jul-Aug, 2026 ISSN: 3049-1835

Macro-Financial Fusion for Next-Session VIX and Credit-Risk-Pressure Direction Forecasting with Evidence-Grounded Risk Memos

Iris Yang

Computer Science, Purdue University, West Lafayette, IN, USA

Received: 25.06.2026 | **Accepted:** 25.07.2026 | **Published:** 27.07.2026

***Corresponding Author:** Iris Yang

DOI: *10.5281/zenodo.21624571*

Abstract **Original Research Article**

Reliable next-session risk monitoring requires models that remain interpretable under limited and mixed-frequency data. This study evaluates a lightweight macro-financial fusion framework for forecasting the direction of the Cboe Volatility Index (VIX) and a transparent credit-risk-pressure index (CRPI) built from eight FRED series. Daily market and Treasury observations are aligned with carried-forward monthly policy, labor, and price indicators; trailing features feed persistence, logistic regression, random forest, gradient boosting, gated recurrent unit, and temporal-transformer models under chronological train-validation-test splits. The VIX task uses 258 observed 2025 rows and yields 162 training, 45 validation, and 41 test sequences after a 10-session window. The longer CRPI task yields 1,614, 502, and 375 sequences after a 21-session window. Random forest provides the strongest VIX class balance (accuracy = 0.634, balanced accuracy = 0.626, F1 = 0.571, ROC-AUC = 0.657), while gradient boosting leads the CRPI task (accuracy = 0.584, balanced accuracy = 0.576, F1 = 0.519, ROC-AUC = 0.604). The compact transformer does not outperform the tree ensembles. A deterministic evidence-grounded memo layer converts probabilities and recorded drivers into eight test-period summaries; complete driver coverage and zero detected numeric violations are consistency checks by construction, not evidence of analyst comprehension or autonomous language-model reasoning. Results support cautious use as a reproducible monitoring and audit-documentation framework. Limitations include the one-year VIX span, the rule-defined CRPI proxy, release-vintage uncertainty, and the absence of human evaluation.

Keywords: VIX, credit-risk pressure, macro-financial forecasting, temporal transformers, evidence-grounded risk communication.

Introduction

Volatility and credit conditions are central to market surveillance, treasury planning, portfolio de-risking, and credit allocation. The Cboe Volatility Index (VIX) provides an observable measure of option-implied equity-market uncertainty, whereas yield-curve, policy-rate, labor-market, price, and equity variables describe complementary dimensions of the macro-financial state (Cboe Exchange, Inc., 2019). Federal Reserve Economic Data offers a

reproducible public source for combining these series, and the FRED-MD tradition demonstrates the value of standardized macroeconomic panels for empirical forecasting (Federal Reserve Bank of St. Louis, n.d.; McCracken & Ng, 2016).

The immediate forecasting target in this study is the sign of the next observed change rather than an exact future level. That choice matches monitoring settings in which the direction of risk matters before a precise point estimate is actionable. Recent work has directly combined news-based uncertainty and macro-market variables for VIX-direction forecasting over a 2015–2024 FRED panel (H. Zhou & Zhang, 2025). The present experiment is narrower in VIX history but broader in its paired design: it evaluates an observed VIX target alongside a transparent credit-risk-pressure proxy and couples both forecasts to the same evidence-record format.

A small FRED export does not contain every variable needed to observe corporate credit risk directly. Accordingly, the second target is a credit-risk-pressure index (CRPI), not a corporate-spread measure. Its construction follows established economic channels: yield-curve flattening can signal weaker future activity (Estrella & Mishkin, 1998); structural credit risk relates firm value to default exposure (Merton, 1974); credit-spread changes reflect macro-financial and liquidity forces (Collin-Dufresne et al., 2001); and equity volatility is associated with corporate bond yields (Campbell & Taksler, 2003). Work on explainable and fair credit scoring similarly emphasizes that a risk score should be accompanied by transparent decision logic rather than treated as an unqualified measure of underlying default risk (Xin, 2026a).

The study also sits within a wider movement toward evidence-grounded financial decision support. Accounting-aware retrieval has been proposed for institutional due diligence on tokenized receivables (Chen et al., 2025), while financial-search research has examined query rewriting, hybrid retrieval, and listwise reranking for returning auditable evidence (Meng et al., 2026). These systems solve different tasks from directional forecasting, but they share a useful design principle: the prediction or recommendation should remain separable from the records used to justify it.

Analyst-facing presentation is therefore part of the research problem rather than an ornamental final step. Financial-risk dashboards have emphasized visual hierarchy, chart comprehension, and explainability for institutional monitoring (Li et al., 2025). Evidence-constrained agents for tokenized-asset monitoring extend the same principle from retrieval to continuing disclosure, settlement, and secondary-market oversight (S. Zhou et al., 2026). These studies motivate a conservative memo layer that reports model facts without claiming independent market judgment.

The modeling comparison bridges econometric time-series practice and compact machine learning. Volatility clustering motivates trailing volatility and momentum features (Bollerslev, 1986; Engle, 1982), while macro demand forecasting illustrates how economic signals and operational response models can be combined without erasing their distinct roles (Lu et al., 2025). Random forests and gradient boosting provide nonlinear tabular benchmarks (Breiman, 2001; Friedman, 2001), and regularized linear estimation provides a stable reference when predictors are correlated (Hoerl & Kennard, 1970).

Sequence models are included as disciplined benchmarks, not presumed winners. The recurrent benchmark follows the gated recurrent unit formulation evaluated by Chung et al. (2014). The transformer uses self-attention (Vaswani et al., 2017) in a compact temporal configuration informed by work on interpretable multi-horizon forecasting (Lim et al., 2021). Training uses AdamW, whose decoupled weight-decay formulation is described by Loshchilov and Hutter (2019).

Probability calibration and visible uncertainty are particularly important when model outputs enter human workflows. Uncertainty-aware late fusion treats confidence calibration as part of the fusion rule rather than an afterthought (Xin, 2025). In other domains, uncertainty-aware medical explanation cards make confidence and supporting visual evidence explicit (Zhong, Wu, et al., 2025), and breast-ultrasound interfaces similarly frame uncertainty as information to communicate rather than conceal (Ye et al., 2025). These are cross-

domain design analogies; they do not validate financial forecasts, but they help motivate the separation of probability, predicted direction, observed drivers, and narrative text.

The memo layer follows the broader language-model practice of converting structured context into readable language (Brown et al., 2020; Devlin et al., 2019), but no autonomous language model is evaluated in this experiment. Instead, a deterministic formatter turns a saved prediction record into an LLM-style memo. Evidence-card work in scientific search emphasizes visible source hierarchy (Jin, 2025b), and accounting-disclosure review cards emphasize links between a claim and its underlying filing evidence (K. Zhang et al., 2025). The present design adopts that traceability goal while restricting every numeric statement to values already stored in the record.

Three questions guide the study. First, can a compact macro-financial panel outperform a directional persistence benchmark for the next observed VIX move? Second, can a rule-defined pressure index support a coherent longer-horizon direction task when direct corporate-spread data are absent? Third, can a forecast record support concise analyst-facing language without introducing unsupported facts? These questions are evaluated with chronological splits, fixed test periods, explicit thresholds, scalar metrics, confusion matrices, regime summaries, and a deterministic memo audit.

The contribution is a reproducible monitoring template rather than a production trading system. It aligns mixed-frequency public data, compares six estimators on two related but distinct targets, and records the probability and observed drivers used in each memo. The design follows the chronological discipline of financial time-series evaluation (Box et al., 2015; Hamilton, 1994; Tsay, 2010) and treats narrative output as documentation of a model signal. It does not claim causal effects, portfolio profitability, point-in-time vintage completeness, or independent language-model reasoning.

Materials and Methods

Data Alignment

The empirical panel is built from three FRED graph-observation files. The first daily file contains VIXCLS and SP500, the second contains DGS10, DGS2, and T10Y2Y, and the monthly file contains FEDFUNDS, UNRATE, and CPIAUCSL. Daily market dates from the VIX and S&P 500 file define the principal calendar. Treasury yields are merged by date and carried forward across intervening missing observations; monthly macroeconomic series are merged by observation date with an as-of rule and carried forward until the next recorded observation. S&P 500 rows with missing market values are excluded from target construction. Observed VIX rows are used only for the VIX task, whereas the longer S&P 500, Treasury, and macro panel supports the CRPI task. This time ordering follows standard forecasting practice for lagged information sets (Box et al., 2015; Hamilton, 1994; Tsay, 2010).

Table I. FRED graph series inventory used in the experiments.

Series	Source file	Frequency	First non-null	Last non-null	Non-null observations	Missing observations
VIXCLS	daily,_close.csv	Daily	2025-01-02	2025-12-31	258	2350
SP500	daily,_close.csv	Daily	2016-07-07	2026-07-06	2512	96

DGS10	daily.csv	Daily	1962-01-02	2026-07-02	16110	720
DGS2	daily.csv	Daily	1976-06-01	2026-07-02	12518	4312
T10Y2Y	daily.csv	Daily	1976-06-01	2026-07-06	12519	4311
FEDFUNDS	monthly.csv	Monthly	1954-07-01	2026-06-01	864	90
UNRATE	monthly.csv	Monthly	1948-01-01	2026-06-01	941	13
CPIAUCSL	monthly.csv	Monthly	1947-01-01	2026-05-01	952	2

The alignment procedure reduces two common sources of leakage. Monthly values are not interpolated between recorded observations, and all rolling features use trailing rather than centered windows. Targets are assigned only after date-t features are fixed, with the next observation used solely to define the label. CRPI standardization uses training-period means and standard deviations that are then frozen for validation and testing. These controls prevent future outcomes and test-period scaling constants from entering model inputs. They do not, however, create a point-in-time vintage database: a FRED observation date can differ from the date on which a release became available and can conceal later revisions. The experiment should therefore be described as chronologically aligned with reduced leakage, not as fully release-vintage clean.

Feature Construction

Market features include S&P 500 one-day log return, 5-, 10-, 21-, and 63-day momentum, realized volatility over the same horizons, and moving-average gaps. Rate features include DGS10, DGS2, T10Y2Y, one-, five-, and 21-day changes, moving averages, the short-minus-long spread, and a curve-level interaction. Macro features include FEDFUNDS, UNRATE, CPIAUCSL, an approximate year-over-year CPI transformation, the real federal funds rate, and trailing changes and averages. VIX features include its level, log level, one-day change, trailing momentum and volatility, and a VIX-S&P 500 divergence measure. Volatility features follow the empirical motivation established by ARCH and GARCH models (Bollerslev, 1986; Engle, 1982).

Table II. Engineered feature groups and representative variables.

Feature group	Count	Examples
Market	13	S&P 500 return, 21-day momentum, realized volatility, moving-average gaps
Rates and curve	17	DGS10, DGS2, T10Y2Y, one-, five-, and 21-day changes
Macro	17	FEDFUNDS, UNRATE, CPI, inflation proxy, real fed funds
VIX state	14	VIX level, one-day change, 5/10/21-day VIX momentum and

		realized volatility
Credit-pressure components	6	curve inversion, fed funds, unemployment, inflation, equity volatility, negative momentum

Feature names remain economically interpretable: momentum describes recent trend, realized volatility describes short-horizon uncertainty, moving-average gaps describe distance from local trend, and rate changes describe curve repricing. The saved data record supplies the three displayed drivers for each selected memo. Although this record-level design is consistent with the broader goal of additive feature explanations (Lundberg & Lee, 2017), the experiment does not independently validate a SHAP-style attribution method or claim that the displayed drivers are causal.

Target Construction

The VIX target equals one when VIXCLS at the next observed VIX date exceeds its current value. CRPI is the equal-weight average of six components standardized with training-period statistics: negative T10Y2Y, FEDFUNDS, UNRATE, year-over-year CPI inflation, S&P 500 21-day realized volatility, and negative S&P 500 21-day momentum. The CRPI target equals one when that constructed index rises at the next market observation. The label therefore represents movement in a transparent macro-financial pressure proxy; it is not a default probability, a corporate-spread forecast, or a firm-level credit score. This distinction is aligned with explainable-risk work that separates a model score from the real-world construct it is intended to support (Xin, 2026a).

The six CRPI components intentionally mix slow and fast variables. FEDFUNDS, UNRATE, and inflation represent slow-moving policy and macroeconomic pressure, whereas the term spread, equity momentum, and realized volatility capture faster market repricing. Equal weighting makes the index reproducible and avoids estimating unstable component weights from a limited panel. This construction is a measurement choice rather than a learned latent-credit model; consequently, feature-target overlap must be considered when interpreting the CRPI results.

Chronological Splits and Sequence Windows

The VIX experiment uses a 10-session window because the observed VIX series spans one calendar year. After feature warm-up and sequence construction, training covers January 15-August 29, 2025; validation covers September 1-October 31, 2025; and testing covers November 3-December 30, 2025. The CRPI experiment uses a 21-session window, with training from August 4, 2016 to December 30, 2022, validation from January 3, 2023 to December 31, 2024, and testing from January 2, 2025 to July 2, 2026. Chronological evaluation is essential for serially dependent data and meaningful forecast comparison (Diebold & Mariano, 1995). Related applied work shows why small-history or changing-regime forecasts should be assessed on later periods rather than randomly mixed rows: few-shot workload forecasting treats new tenants as a cold-start problem (He et al., 2025), while transformer-based bike-demand forecasting explicitly considers changing weather and seasonal regimes (Xin, 2026b). These are evaluation analogies, not evidence about the financial scores reported here.

Table III. Chronological split design and class balance.

Task	Split	Date start	Date end	Rows	Up class rate	Sequence length
VIX direction	train	2025-01-15	2025-08-29	162	0.438	10
VIX direction	val	2025-09-01	2025-10-31	45	0.489	10
VIX direction	test	2025-11-03	2025-12-30	41	0.439	10
Credit-risk pressure direction	train	2016-08-04	2022-12-30	1614	0.403	21
Credit-risk pressure direction	val	2023-01-03	2024-12-31	502	0.456	21
Credit-risk pressure direction	test	2025-01-02	2026-07-02	375	0.445	21

Models and Training

Six estimators are evaluated. Persistence maps the previous realized direction to probabilities of 0.65 and 0.35. Logistic regression uses median imputation, z-score scaling, class weights, and C = 0.5, providing a regularized linear reference (Hoerl & Kennard, 1970). Random forest uses 400 trees, maximum depth 5, minimum leaf size 5, and balanced subsampling (Breiman, 2001); gradient boosting uses 160 estimators, learning rate 0.035, depth 2, and subsample 0.75 (Friedman, 2001). The GRU contains one recurrent layer with 24 hidden units (Chung et al., 2014). The temporal transformer contains one encoder layer, model dimension 24, four attention heads, dropout 0.12, and feed-forward width 48 (Lim et al., 2021; Vaswani et al., 2017). Neural models use class-weighted binary cross-entropy, AdamW, and early stopping; decoupled weight decay distinguishes the optimizer applied here (Loshchilov & Hutter, 2019).

Table IV. Model configurations used in the forecasting comparisons.

Model	Configuration
Persistence	Previous realized direction converted to probabilities 0.65/0.35; validation balanced-accuracy threshold.
Logistic regression	Median imputation, z-score scaling, class-weighted lbfgs logistic regression, C=0.5.
Random forest	400 trees, maximum depth 5, minimum leaf size 5,

	balanced subsampling.
Gradient boosting	160 estimators, learning rate 0.035, depth 2, subsample 0.75.
GRU	One recurrent layer, 24 hidden units, AdamW, class-weighted BCE, early stopping on validation loss.
Temporal transformer	One encoder layer, d_model=24, four attention heads, AdamW, class-weighted BCE, early stopping.

Evaluation Metrics and Thresholds

The reported table metrics are accuracy, balanced accuracy, precision, recall, F1, ROC-AUC, Brier score, and the validation-selected threshold. Balanced accuracy is primary because each task is near, but not exactly at, equal prevalence; precision, recall, and F1 clarify upward-risk detection, ROC-AUC measures ranking, and Brier score summarizes squared probability error (Powers, 2011). Thresholds maximize validation balanced accuracy and are then fixed for testing. Confidence calibration is treated as part of operational fusion rather than cosmetic post-processing (Xin, 2025), but the sample is too small to support formal calibration guarantees, especially for the 41-observation VIX test. Cross-domain explanation cards make confidence visible to reviewers (Ye et al., 2025; Zhong, Wu, et al., 2025); they motivate separating the probability from its verbal interpretation but do not validate the financial model.

The implementation uses a fixed random seed of 271828 for ensemble initialization, neural initialization, and mini-batch ordering. Classical models receive the current-day vector; GRU and transformer models receive trailing sequences. Missing values are imputed with training medians, and neural features are scaled with training-sequence parameters only. Validation selects stopping points and thresholds, after which test periods remain fixed. Offline uplift evaluation under privacy constraints similarly distinguishes policy selection from final evaluation (Bai et al., 2026), while multi-horizon GPU-demand studies link forecasts to operational risk curves (Zhao et al., 2025). These studies concern different domains and serve only as precedents for separating estimation, policy selection, and held-out assessment.

Evidence-Grounded Memo Layer

Each selected memo receives a structured record containing the task, date, model probability, predicted direction, realized direction, and three recorded drivers. A deterministic template then produces a short LLM-style paragraph constrained to those fields. The architecture echoes evidence-centered pipelines in which retrieval is bounded by an explicit budget (Su et al., 2025), provenance is preserved across an evidence chain (Jin, 2025a), and unanswerability is evaluated against retrieval coverage (Zhong, Chen, et al., 2025). The forecast, action threshold, and communication layer remain separate. Operational studies make similar distinctions when probabilities inform admission-control policy memos (Zhao et al., 2026) or workload forecasts feed power-aware inventory planning (He et al., 2026). Narrative-aware verification further motivates checking generated claims against ranked evidence (Su, Chen, et al., 2026). Here, there is no retrieval corpus or autonomous generator: direction agreement and driver coverage are consistency checks by construction, and numeric checking is limited to the saved record.

Validation data are used only for threshold selection and neural early stopping; final scores are computed on fixed test intervals. Metrics, predictions, figures, training histories, and memo records are stored separately so that a reviewer can inspect a model score without accepting its narrative rendering. LLM-compatible infrastructure dashboards similarly

translate forecast risk into reviewable visual briefs (Liu et al., 2025), but this study evaluates only textual summaries and makes no claim that the memo improves human decisions.

Results and Discussion

Data and Target Characterization

Table I shows the raw inventory. VIX has 258 non-null observations from January 2 through December 31, 2025; S&P 500 has 2,512 non-null observations from July 7, 2016 through July 6, 2026. Treasury histories are longer, but the merged experiments follow the market calendar. FEDFUNDS, UNRATE, and CPIAUCSL act as monthly state variables. The paired tasks are not symmetric: VIX is observed, whereas CRPI is constructed from variables that also appear among predictors. Numerical guardrails for research assistants provide a useful governance analogy—computed values should remain reproducible from named data fields and bounded by explicit checks (Li et al., 2026)—but they do not resolve the measurement limits of CRPI.

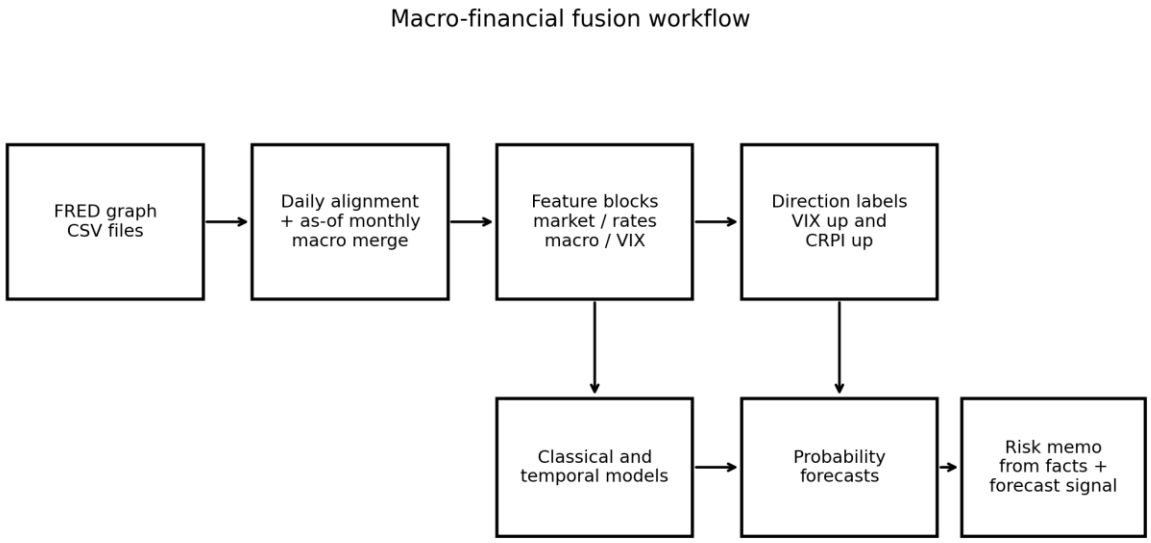

Figure 1. Macro-financial fusion workflow from FRED graph observations to forecasts and risk memos.

Figure 2 plots normalized VIX and S&P 500 values over the observed VIX sample. The graph shows that the VIX target has meaningful daily variation even though it covers only one year. Figure 3 plots the CRPI over the longer panel. The index moves around zero because each component is standardized on the training period, and it rises when curve inversion, high policy rates, inflation pressure, labor-market pressure, equity volatility, or negative equity momentum become stronger. Figure 4 reports the correlation structure among selected variables and confirms that the panel is neither purely market-driven nor purely macro-driven; equity volatility, rates, inflation, and labor variables capture different portions of the risk state.

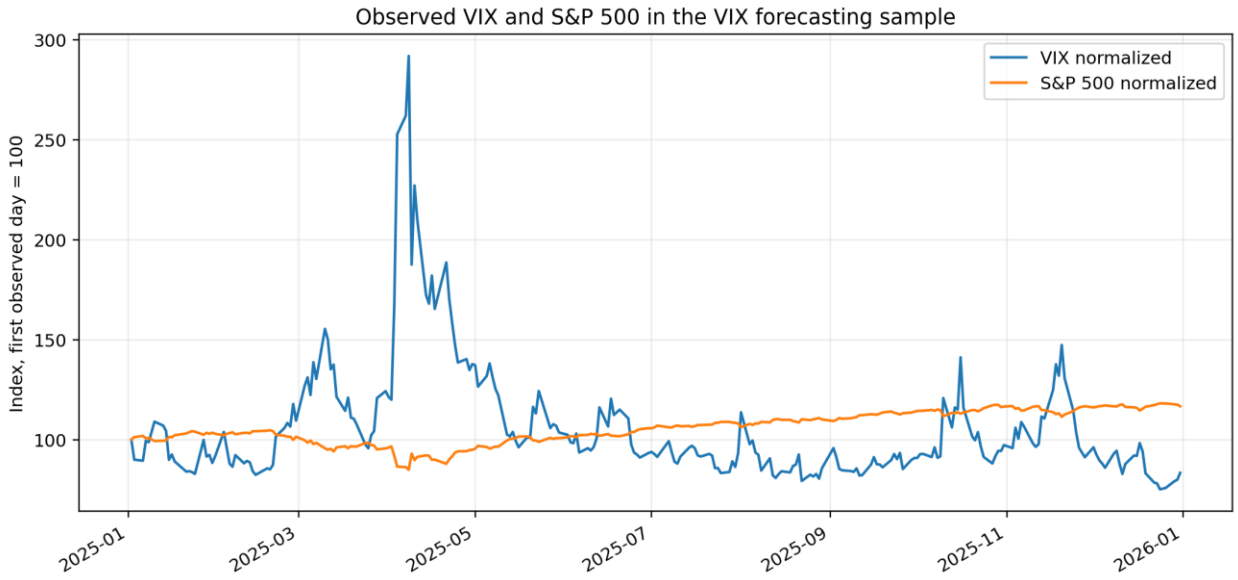

Figure 2. Normalized VIX and S&P 500 values in the observed 2025 VIX sample.

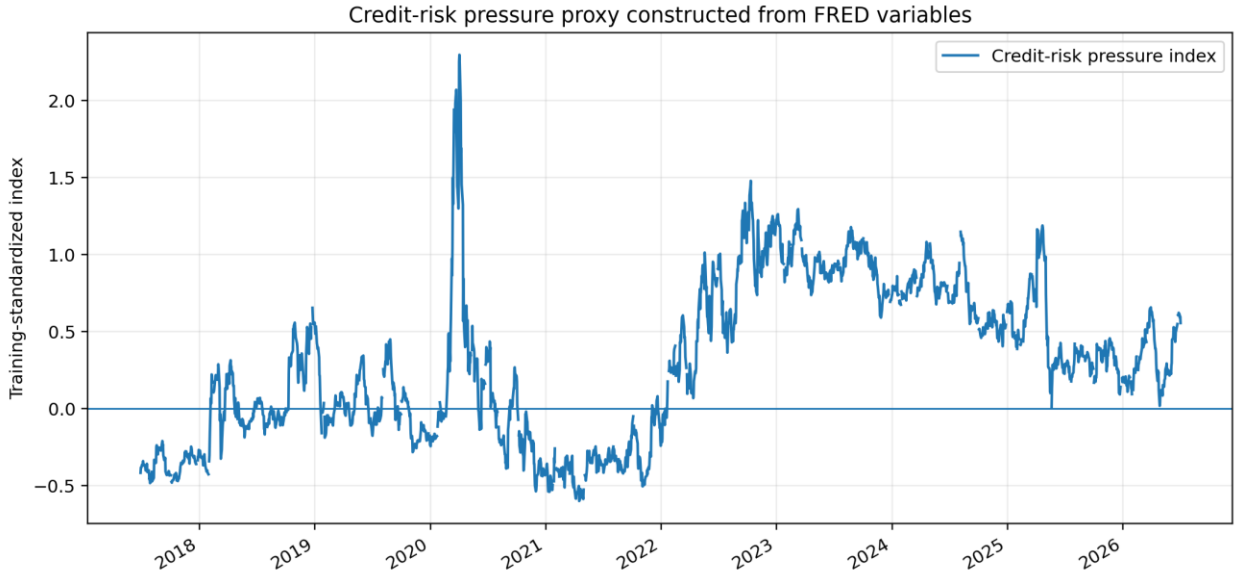

Figure 3. Credit-risk pressure index constructed from the available FRED variables.

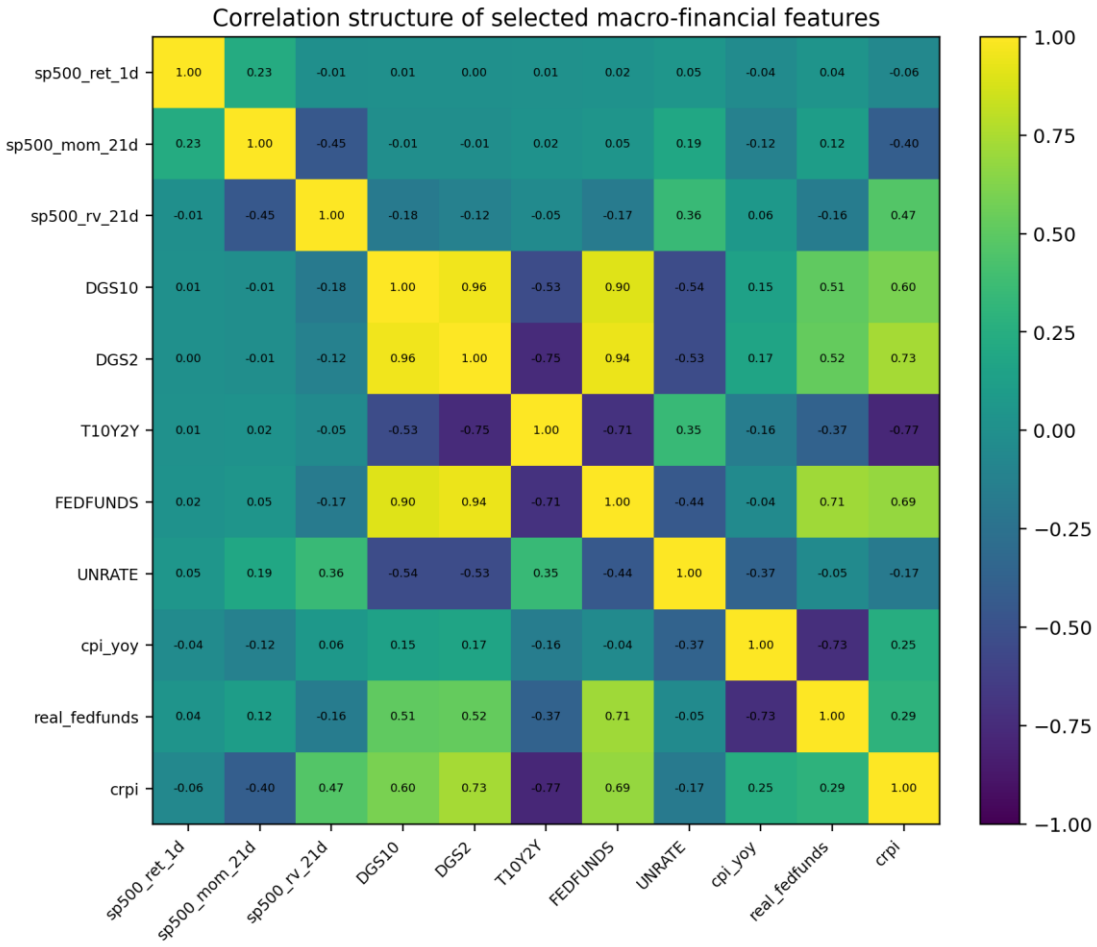

Figure 4. Correlation heatmap for selected macro-financial variables and CRPI.

The split design also produces two complementary tests. The VIX test is short but direct: the observed target is an established volatility index, and the question is whether recent macro-financial information helps classify the next observed move. The credit-pressure test is longer but constructed: the target is the next movement in a transparent composite index. Treating these tests together is useful because it separates target observability from sample length. The VIX task emphasizes direct market meaning under limited history, while the CRPI task emphasizes longer-run macro-financial state tracking under a proxy target.

VIX Direction Results

The VIX comparison is reported in Table V and Figure 5. Random forest provides the strongest balanced classification, with accuracy of 0.634, balanced accuracy of 0.626, F1 of 0.571, ROC-AUC of 0.657, and Brier score of 0.239. The GRU is close on class balance and has the highest VIX ROC-AUC, 0.671, whereas logistic regression attains F1 of 0.643 by predicting every test observation as upward at its selected threshold. The temporal transformer reaches 0.561 accuracy and 0.518 balanced accuracy. Thus, self-attention does not outperform simpler models in this one-year sample. H. Zhou and Zhang (2025) provide a direct comparison by studying news uncertainty and macro-market fusion on a longer 2015–2024 VIX panel, but differences in dates, features, splits, and models preclude a numerical superiority claim.

Table V. VIX direction forecasting results on the chronological test set.

Model	Accuracy	Bal. acc.	Precision	Recall	F1	ROC-AUC	Brier	Threshold
Random forest	0.634	0.626	0.588	0.556	0.571	0.657	0.239	0.535
GRU	0.610	0.616	0.545	0.667	0.600	0.671	0.251	0.585
Logistic regression	0.512	0.565	0.474	1.000	0.643	0.630	0.439	0.725
Gradient boosting	0.537	0.551	0.480	0.667	0.558	0.662	0.242	0.445
Temporal transformer	0.561	0.518	0.500	0.167	0.250	0.551	0.260	0.600
Persistence	0.439	0.500	0.439	1.000	0.610	0.477	0.276	0.200

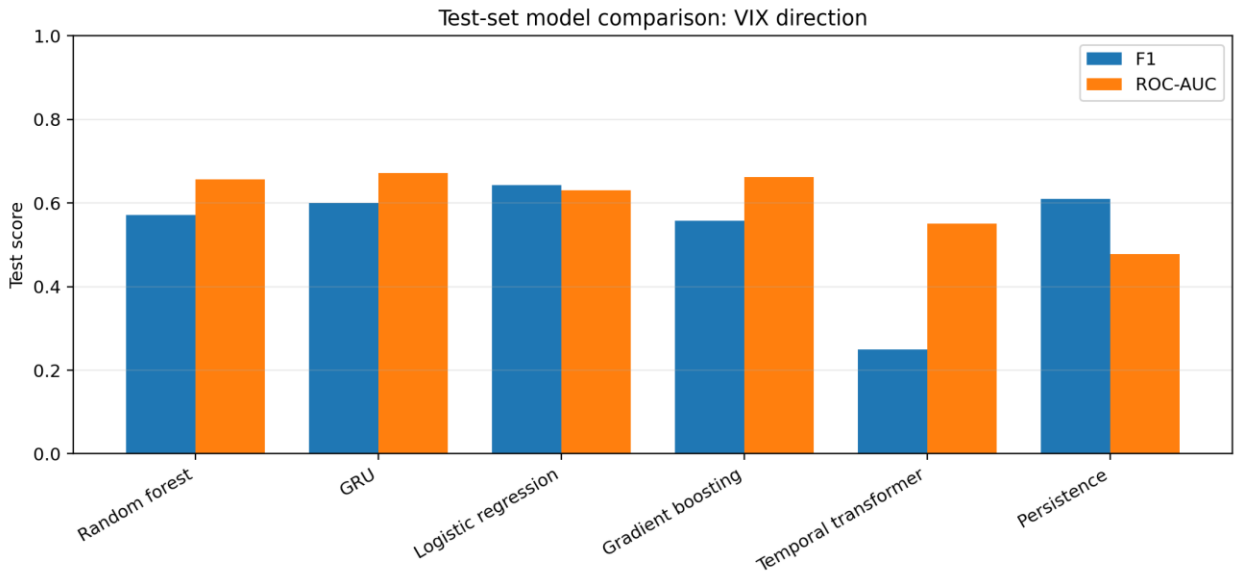

Figure 5. VIX direction model comparison by F1 and ROC-AUC.

The VIX table also exposes the limitation of persistence. Its validation-selected threshold of 0.200 lies below both mapped probabilities, 0.35 and 0.65, so every test observation is classified upward. That yields recall of 1.000 and F1 of 0.610 but balanced accuracy of exactly 0.500. Random forest therefore improves two-sided discrimination rather than merely reproducing an always-up alert. Because the

VIX test contains only 41 observations, the difference is descriptive and should not be interpreted as a stable multi-cycle edge.

Credit-Risk-Pressure Direction Results

The credit-risk pressure comparison is reported in Table VI and visualized in Figure 6. Gradient boosting is the best model on balanced accuracy, with 0.584 accuracy, 0.576 balanced accuracy, 0.519 F1, 0.604 ROC-AUC, and 0.250 Brier score. Persistence and logistic regression both classify nearly all test observations as upward pressure at the selected threshold, producing 0.445 accuracy and 0.500 balanced accuracy. Random forest has 0.555 accuracy and the lowest Brier score among the main credit models, 0.249, but its recall is low at the selected threshold. The GRU and temporal transformer are less effective on the credit task, with 0.507 and 0.502 balanced accuracy respectively. The tree ensembles handle the compact tabular structure more effectively than the neural sequence models in this panel.

Table VI. Credit-risk pressure direction forecasting results on the chronological test set.

Model	Accuracy	Bal. acc.	Precision	Recall	F1	ROC-AUC	Brier	Threshold
Gradient boosting	0.584	0.576	0.535	0.503	0.519	0.604	0.250	0.560
Random forest	0.555	0.507	0.500	0.072	0.126	0.569	0.249	0.580
GRU	0.512	0.507	0.453	0.461	0.457	0.506	0.291	0.660
Temporal transformer	0.448	0.502	0.446	0.994	0.616	0.564	0.403	0.700
Logistic regression	0.445	0.500	0.445	1.000	0.616	0.594	0.554	0.795
Persistence	0.445	0.500	0.445	1.000	0.616	0.469	0.280	0.200

Figure 6. Credit-risk pressure direction model comparison by F1 and ROC-AUC.

Thresholds, Calibration, and Error Structure

Threshold behavior is informative. Logistic regression achieves high recall on both tasks by assigning nearly every test observation to the upward class, while its Brier scores indicate poor probability quality. Random forest produces the best VIX balance, and gradient boosting produces the best CRPI balance by combining market, rate, and macro states with shallow nonlinear interactions. The neural models are not ruled out; the experiment shows only that added sequence capacity does not translate into higher balanced accuracy under these samples. Cross-domain forecasting studies likewise treat complexity as contingent on data and operational loss rather than as an end in itself (He et al., 2025; Zhao et al., 2026).

Feature-Fusion Ablation

Table VII reports feature-fusion ablations. For VIX direction, full fusion with gradient boosting gives the highest balanced accuracy among the tested feature sets, 0.551, and ROC-AUC of 0.662. Rates alone have the same ROC-AUC but lower balanced accuracy, and VIX-state features alone yield 0.547 balanced accuracy. The 27-variable row originally labeled “Market only” combines the 13 market and 14 VIX-state variables and should therefore be read as “Market + VIX state.” For CRPI, full fusion reaches 0.521 balanced accuracy, only slightly above equity-only and market-plus-rates variants. No block is universally sufficient across both tasks.

Table VII. Feature-fusion ablation results using gradient boosting.

Task	Feature set	Count	Accuracy	Bal. acc.	F1	ROC-AUC	Brier
VIX direction	VIX state only	14	0.512	0.547	0.600	0.609	0.253

VIX direction	Market + VIX state	27	0.488	0.543	0.632	0.645	0.245
VIX direction	Rates only	17	0.463	0.522	0.621	0.662	0.311
VIX direction	Macro only	17	0.439	0.500	0.610	0.467	0.321
VIX direction	Full fusion	61	0.537	0.551	0.558	0.662	0.242
Credit-risk pressure direction	Equity only	13	0.469	0.520	0.622	0.617	0.240
Credit-risk pressure direction	Rates only	17	0.448	0.502	0.617	0.466	0.259
Credit-risk pressure direction	Macro only	17	0.445	0.500	0.616	0.491	0.258
Credit-risk pressure direction	Market + rates	30	0.451	0.504	0.616	0.616	0.251
Credit-risk pressure direction	Full fusion	47	0.472	0.521	0.621	0.604	0.250

The ablation table also explains why the model comparison should not be reduced to a single headline score. VIX-state features are informative, but market and rate features change the probability ranking. Credit pressure is naturally tied to equity variables because CRPI contains realized volatility and negative momentum, yet the full feature set adds rate and macro context that slightly improves balanced accuracy. Macro-only models remain weak at the daily horizon because FEDFUNDS, UNRATE, and CPI move slowly relative to the target. The results therefore support a fusion design in which low-frequency macro variables set the state, while daily market variables provide the short-horizon timing signal.

Figure 7 displays the best-model confusion matrices. The VIX random forest correctly classifies 16 downward and 10 upward observations in the 41-row test set, with 7 false upward forecasts and 8 missed upward forecasts. The credit gradient boosting model correctly classifies 135 downward and 84 upward observations in the 375-row test set, with 73 false upward forecasts and 83 missed upward forecasts. These matrices clarify why balanced accuracy is more informative than F1 alone: a model can obtain high F1 by leaning heavily toward the upward class, while balanced accuracy exposes whether both directions are being recognized.

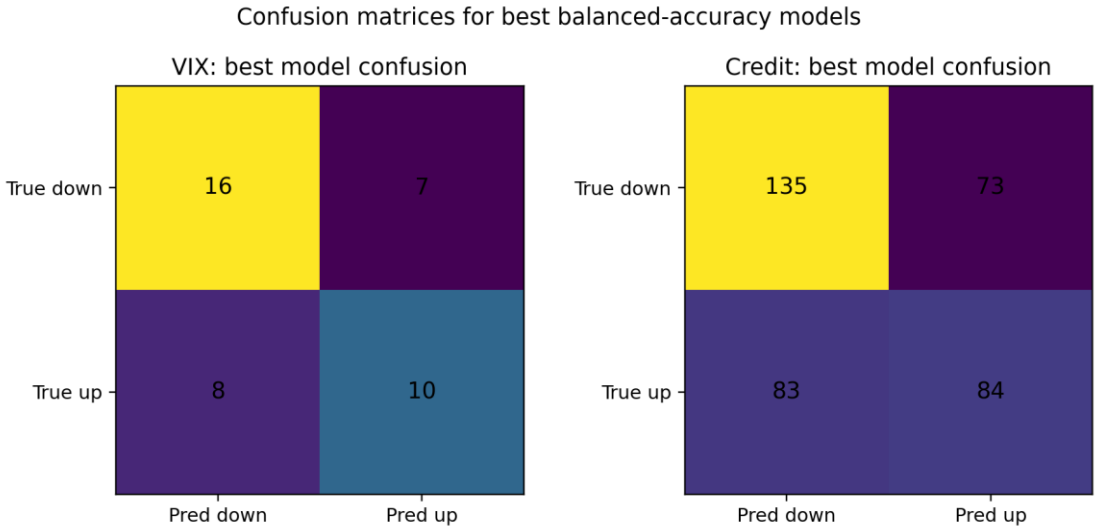

Figure 7. Confusion matrices for the best balanced-accuracy models.

Regime-Specific Performance

Table VIII contains the regime subsets preserved in the saved export. In the VIX test, random forest accuracy is 0.625 in low VIX, 0.630 in medium VIX, and 0.667 in high VIX; the high-VIX subset has only six observations and F1 of zero. For CRPI, the displayed middle- and high-pressure subsets contain 276 and 82 observations. They total 358 rather than the 375 test rows in Table III, leaving 17 low-pressure observations absent from the saved regime table. No missing metrics are reconstructed. The CRPI regime discussion therefore describes only the exported subsets. Cross-domain studies model weather, seasonal, and workload regimes explicitly (Xin, 2026b; Zhao et al., 2025), but the current smaller and incomplete table supports only descriptive heterogeneity.

Table VIII. Best-model performance for the risk-regime subsets preserved in the saved output.

Task	Regime	N	Accuracy	F1	Mean P(up)	Actual up rate
VIX direction	low VIX	8	0.625	0.769	0.688	0.625
VIX direction	medium VIX	27	0.630	0.500	0.520	0.407
VIX direction	high VIX	6	0.667	0.000	0.432	0.333
Credit-risk pressure direction	middle pressure	276	0.565	0.574	0.572	0.482

Credit-risk pressure direction	high pressure	82	0.561	0.143	0.495	0.415

Evidence-Grounded Memo Evaluation

Table IX evaluates eight selected test-period memos. Mean length is 74.125 words, driver coverage is 1.000, direction agreement is 1.000, detected numeric violations are zero, and prediction accuracy on selected memo days is 0.750. The first three results are consistency checks of a deterministic template: drivers and model direction are inserted from the saved record, and numeric statements are checked against it. They are not independent measures of explanation quality, human comprehension, factual completeness, or autonomous language-model reasoning. Table X displays four VIX examples only; the other four audited memo records are not shown and should not be described as CRPI excerpts.

Table IX. Risk-memo layer evaluation.

Metric	Value
Memo count	8.000
Mean word count	74.125
Driver coverage	1.000
Direction agreement	1.000
Factual violations	0.000
Prediction accuracy on memo days	0.750

Table X. Representative VIX risk-memo excerpts.

Task	Date	Predicted direction	Probability of up	Actual direction	Correct	Memo summary
VIX	2025-12-24	Up	0.755	Up	1	Risk memo for 2025-12-24: the VIX direction model assigns a 75.5% probability to an upward

						next-session move, so the for...
VIX	2025-12-26	Up	0.731	Up	1	Risk memo for 2025-12-26: the VIX direction model assigns a 73.1% probability to an upward next-session move, so the for...
VIX	2025-11-17	Down	0.358	Up	0	Risk memo for 2025-11-17: the VIX direction model assigns a 35.8% probability to an upward next-session move, so the for...
VIX	2025-11-18	Down	0.361	Down	1	Risk memo for 2025-11-18: the VIX direction model assigns a 36.1% probability to an upward next-session move, so the for...

The displayed VIX examples illustrate how a probability, predicted direction, and recorded drivers can be compressed into a readable note. Evidence-card research in e-commerce similarly separates a recommendation from evidence used to calibrate trust (B. Zhang et al., 2025), while scientific-search interfaces emphasize visible evidence hierarchy (Jin, 2025b). These precedents support the presentation

choice, not forecast accuracy. Because no analyst study was conducted, the experiment cannot determine whether the memos improve decisions, reduce reading time, or create automation bias.

For an applied risk reader, the outputs are monitoring signals rather than automatic decisions. Financial-risk dashboards emphasize that hierarchy and chart comprehension shape how institutional evidence is read (Li et al., 2025), and infrastructure visual briefs translate forecast risk into reviewable operational cues (Liu et al., 2025). Here, an upward probability means only that the fitted model associates the current feature record with a higher next-observation target. The memo documents that association; it does not establish causality, recommend a trade, or replace analyst review.

Operational Interpretation

Overall, a compact FRED panel contains measurable directional signal, yet model complexity must match sample size and target construction. Tree ensembles lead the final balanced-accuracy comparisons, while the compact transformer serves as a sequence benchmark. The memo is most defensible when constrained to recorded facts. Incident-visualization work highlights evidence-constrained, reviewer-editable summaries (Nie et al., 2026), and DevOps copilots emphasize provenance and command safety before automated action (B. Zhang et al., 2026). The analogous governance implication is to require source inspection and human confirmation before a signal enters a risk workflow.

Calibration evidence also affects use. The VIX random forest combines the strongest balanced accuracy with a Brier score of 0.239, while the CRPI gradient booster trades a slightly higher Brier score than random forest for substantially better two-sided classification. This does not establish perfect calibration. It supports balanced accuracy as the selection criterion and Brier score as a secondary probability-quality check. Calibration-aware fusion reinforces the need to examine confidence before combining signals (Xin, 2025), and privacy-robust policy evaluation illustrates why a selected decision rule should be assessed on fixed holdout data (Bai et al., 2026).

The regime analysis reinforces the same point. Performance changes across low, middle, and high-risk states because the empirical meaning of an upward move depends on the level from which it starts. A small VIX increase from a calm state can be easier to classify than an increase from an already elevated state, where reversals and mean reversion become more common. Credit pressure shows a similar pattern: once the CRPI is already high, the next movement can be dominated by short-term market noise rather than by the slowly moving macro components. Reporting regime results prevents the paper from overstating the average test score and gives risk users a clearer view of where the system is more reliable.

The memo layer is evaluated on selected high- and low-signal days, so the 0.750 memo-day accuracy is not an unbiased estimate for every test date. The audit rewards field coverage, direction consistency, and numeric traceability rather than eloquence. Narrative-aware verification checks claims against ranked evidence (Su, Chen, et al., 2026). The deterministic formatter avoids agentic risks but provides none of an autonomous model's open-ended reasoning. Accounting-aware monitoring preserves links across evidence states (S. Zhou et al., 2026), and financial-search reranking preserves query and ranking provenance (Meng et al., 2026); analogously, a production memo should expose source series, timestamps, transformations, model version, and threshold.

Limitations and Scope

The first limitation is sample and measurement scope. The VIX file contains one year of non-null observations, leaving 41 test sequences after feature warm-up and the 10-session window. Results are descriptive of late 2025 rather than evidence across volatility cycles; no confidence intervals, repeated-origin tests, or formal significance comparisons are available. CRPI is an equal-weight pressure proxy built from six observed components, not a corporate-spread series, default outcome, or firm-level credit measure. Because several components also appear as predictors, part of the task reflects persistence and interaction within the constructed index. Future work

should test direct spread, default, and balance-sheet outcomes and compare alternative weighting rules.

The second limitation concerns timing, model validation, and probability assessment. Observation-date as-of merging and trailing windows reduce leakage, but the files are not a vintage-aware release database; publication lags and later revisions may make some monthly values unavailable on their nominal observation dates. The compact transformer also cannot represent a test of large attention architectures. Thresholds are selected on one validation interval, and Brier scores are reported without formal recalibration or interval estimates. Stronger work should use real-time vintages, rolling origins, reliability diagrams, bootstrap or conformal uncertainty, and refusal rules for weak evidence, consistent with calibration and numerical-guardrail principles (Li et al., 2026; Xin, 2025).

The final limitation is narrative and decision scope. Driver coverage, direction agreement, and zero detected numeric violations follow largely from deterministic construction, and no human study measures comprehension, calibration of trust, usefulness, or automation bias. Evidence-constrained financial agents and privacy-risk interfaces show why provenance, secure oversight, and correction mechanisms matter when text enters a decision process (Su, Rao, et al., 2026; S. Zhou et al., 2026). The forecasts are not allocations, hedges, or trading advice; no transaction costs, position limits, utility function, or economic-value backtest is included. Production use would require source-level audit trails, drift and model-version controls, and explicit human approval, as emphasized by reviewer-oriented incident and DevOps interfaces (Nie et al., 2026; B. Zhang et al., 2026).

Concluding Implications

This study presents a complete macro-financial fusion experiment for next-observation VIX and CRPI direction. The panel combines eight FRED series, trailing market, rate, macro, and VIX features, chronological train-validation-test splits, six estimators, and a deterministic evidence-record layer. Random forest is the strongest VIX classifier by balanced accuracy (0.626), and gradient boosting is strongest for CRPI (0.576). These modest, sample-specific scores show that a compact public panel can support a reproducible monitoring experiment. They also show why model, threshold, and communication choices must remain separate: the transformer does not dominate the tree ensembles, and high F1 can arise when nearly every observation is classified upward.

The most defensible deployment path is a monitored analyst aid. Preserve probability and threshold, expose observations and transformations, generate only record-grounded text, and require review before action. Evidence-chain and retrieval-card research motivates maintaining provenance (Jin, 2025a; Zhong, Chen, et al., 2025), while accounting-review cards keep supporting records visible to human reviewers (K. Zhang et al., 2025). Future work should add longer VIX history, real-time macro vintages, direct credit outcomes, rolling-origin validation, formal calibration assessment, and human evaluation. Those extensions can strengthen the system without obscuring the boundary between data, forecast, explanation, and decision.

References

Bai, J., Wang, H., Wu, Q., & Zhang, B. (2026). Privacy-robust incrementality estimation in cookieless settings via uplift modeling: Reproducible evidence from the Hillstrom E-Mail Experiment. *Journal of Technology Informatics and Engineering, 5*(1), 17–38. https://doi.org/10.51903/jtie.v5i1.468

Bollerslev, T. (1986). Generalized autoregressive conditional heteroskedasticity. *Journal of Econometrics, 31*(3), 307–327. https://doi.org/10.1016/0304-4076(86)90063-1

Box, G. E. P., Jenkins, G. M., Reinsel, G. C., & Ljung, G. M. (2015). *Time series analysis: Forecasting and control* (5th ed.). Wiley.

Breiman, L. (2001). Random forests. *Machine Learning, 45*(1), 5–32. https://doi.org/10.1023/A:1010933404324

Brown, T. B., Mann, B., Ryder, N., Subbiah, M., Kaplan, J., Dhariwal, P., Neelakantan, A.,

Shyam, P., Sastry, G., Askell, A., Agarwal, S., Herbert-Voss, A., Krueger, G., Henighan, T., Child, R., Ramesh, A., Ziegler, D. M., Wu, J., Winter, C., . . . Amodei, D. (2020). Language models are few-shot learners. In *Advances in Neural Information Processing Systems, 33,* 1877–1901. https://proceedings.neurips.cc/paper/2020/hash/1457c0d6bfcb4967418bfb8ac142f64a-Abstract.html

Campbell, J. Y., & Taksler, G. B. (2003). Equity volatility and corporate bond yields. *The Journal of Finance, 58*(6), 2321–2350. https://doi.org/10.1046/j.1540-6261.2003.00607.x

Cboe Exchange, Inc. (2019). *Cboe Volatility Index white paper*. https://www.cboe.com/tradable_products/vix/vix_white_paper/

Chen, Y., Zhou, S., & Lin, E. (2025). Accounting-aware evidence retrieval for institutional due diligence of tokenized trade receivable RWA. *Journal of Technology Informatics and Engineering, 4*(3), 649–663. https://doi.org/10.51903/jtie.v4i3.542

Chung, J., Gulcehre, C., Cho, K., & Bengio, Y. (2014). *Empirical evaluation of gated recurrent neural networks on sequence modeling* [Preprint]. arXiv. https://doi.org/10.48550/arXiv.1412.3555

Collin-Dufresne, P., Goldstein, R. S., & Martin, J. S. (2001). The determinants of credit spread changes. *The Journal of Finance, 56*(6), 2177–2207. https://doi.org/10.1111/0022-1082.00387

Devlin, J., Chang, M.-W., Lee, K., & Toutanova, K. (2019). BERT: Pre-training of deep bidirectional transformers for language understanding. In *Proceedings of the 2019 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies,* 4171–4186. https://doi.org/10.18653/v1/N19-1423

Diebold, F. X., & Mariano, R. S. (1995). Comparing predictive accuracy. *Journal of Business & Economic Statistics, 13*(3), 253–263. https://doi.org/10.1080/07350015.1995.10524599

Engle, R. F. (1982). Autoregressive conditional heteroscedasticity with estimates of the variance of United Kingdom inflation. *Econometrica, 50*(4), 987–1007. https://doi.org/10.2307/1912773

Estrella, A., & Mishkin, F. S. (1998). Predicting U.S. recessions: Financial variables as leading indicators. *The Review of Economics and Statistics, 80*(1), 45–61. https://doi.org/10.1162/003465398557320

Federal Reserve Bank of St. Louis. (n.d.). *Federal Reserve Economic Data (FRED)* [Database]. Retrieved July 23, 2026, from https://fred.stlouisfed.org/

Friedman, J. H. (2001). Greedy function approximation: A gradient boosting machine. *The Annals of Statistics, 29*(5), 1189–1232. https://doi.org/10.1214/aos/1013203451

Hamilton, J. D. (1994). *Time series analysis*. Princeton University Press.

He, S., Li, C., & Rao, H. (2025). Few-shot cold-start workload forecasting for new AI inference tenants with time-series foundation models. *Journal of Technology Informatics and Engineering, 4*(1), 306–324. https://doi.org/10.51903/jtie.v4i1.546

He, S., Nie, J., & Li, C. (2026). Power-aware inventory planning for AI infrastructure using job-level forecasting and LLM workload explanations. *Journal of Technology Informatics and Engineering, 5*(1), 341–359. https://doi.org/10.51903/jtie.v5i1.548

Hoerl, A. E., & Kennard, R. W. (1970). Ridge regression: Biased estimation for nonorthogonal problems. *Technometrics, 12*(1), 55–67. https://doi.org/10.1080/00401706.1970.10488634

Jin, J. (2025a). Evidence-chain reliable RAG: Hallucination detection, source attribution, and deterministic provenance explanations. *Journal of Technology Informatics and Engineering, 4*(2), 520–533.

https://doi.org/10.51903/jtie.v4i2.535

Jin, J. (2025b). LLM-style evidence cards for scientific search interfaces: A UI/UX design framework for retrieval transparency, ranking trust, and visual evidence hierarchy. *International Journal of Graphic Design, 3*(2), 397–414. https://doi.org/10.51903/ijgd.v3i2.3698

Li, Z., Zhang, K., & Wong, A. (2026). Numerical-reasoning guardrails for a quant research assistant: A compact reproducible benchmark using SEC and FRED data. *Journal of Technology Informatics and Engineering, 5*(2), 75–90. https://doi.org/10.51903/jtie.v5i2.541

Li, Z., Zhou, S., & Zhou, Z. (2025). Financial risk dashboard design for institutional RWA investors: Visual hierarchy, chart comprehension, and explainability in FinChart-Bench. *International Journal of Graphic Design, 3*(1), 196–210. https://doi.org/10.51903/ijgd.v3i1.3715

Lim, B., Arık, S. Ö., Loeff, N., & Pfister, T. (2021). Temporal fusion transformers for interpretable multi-horizon time series forecasting. *International Journal of Forecasting, 37*(4), 1748–1764. https://doi.org/10.1016/j.ijforecast.2021.03.012

Liu, G., He, S., & Wong, H. (2025). LLM-compatible visual brief cards for AI infrastructure capacity dashboards: A UI/UX framework for turning forecast risk into graphic design decisions. *International Journal of Graphic Design, 3*(1), 196–213. https://doi.org/10.51903/ijgd.v3i1.3723

Loshchilov, I., & Hutter, F. (2019). Decoupled weight decay regularization. In *7th International Conference on Learning Representations (ICLR 2019)*. https://openreview.net/forum?id=Bkg6RiCqY7

Lu, Y., Zhou, H., & Zhang, Y. (2025). A constrained, data-driven budgeting framework integrating macro demand forecasting and marketing response modeling. *Journal of Technology Informatics and Engineering, 4*(3), 493–520. https://doi.org/10.51903/jtie.v4i3.466

Lundberg, S. M., & Lee, S.-I. (2017). A unified approach to interpreting model predictions. In *Advances in Neural Information Processing Systems, 30*, 4765–4774. https://proceedings.neurips.cc/paper/7062-a-unified-approach-to-interpreting-model-predictions

McCracken, M. W., & Ng, S. (2016). FRED-MD: A monthly database for macroeconomic research. *Journal of Business & Economic Statistics, 34*(4), 574–589. https://doi.org/10.1080/07350015.2015.1086655

Meng, S., Chen, J., & Zheng, I. (2026). LLM-inspired offline reranking for financial search: Query rewriting, hybrid retrieval, and listwise relevance ranking on FiQA. *Journal of Technology Informatics and Engineering, 5*(1), 361–378. https://doi.org/10.51903/jtie.v5i1.537

Merton, R. C. (1974). On the pricing of corporate debt: The risk structure of interest rates. *The Journal of Finance, 29*(2), 449–470. https://doi.org/10.1111/j.1540-6261.1974.tb03058.x

Nie, J., Liu, G., Li, C., & Zou, T. (2026). Evidence-constrained incident visualization cards for distributed cloud logs: A UI/UX framework for turning Hadoop, OpenStack, and ZooKeeper logs into actionable SRE design interfaces. *International Journal of Graphic Design, 4*(1), 179–185. https://doi.org/10.51903/ijgd.v4i1.3703

Powers, D. M. W. (2011). Evaluation: From precision, recall and F-measure to ROC, informedness, markedness and correlation. *Journal of Machine Learning Technologies, 2*(1), 37–63.

Su, W., Chen, S., & Qian, E. (2026). Narrative-aware scientific claim verification agent with evidence ranking for ClimateCheck. *Journal of Technology Informatics and Engineering, 5*(1), 327–340. https://doi.org/10.51903/jtie.v5i1.549

Su, W., Chen, S., & Zhao, C. (2025). Budgeted multi-hop retrieval agent for compositional question answering: A retrieval-policy evaluation on the official MultiHop-RAG

benchmark. *Journal of Technology Informatics and Engineering, 4*(3), 649–662. https://doi.org/10.51903/jtie.v4i3.543

Su, W., Rao, H., & Ma, E. (2026). Privacy and data-integrity risk cards for LLM agents: A UI/UX design framework for secure human oversight under prompt-injection attacks. *International Journal of Graphic Design, 4*(1), 186–191. https://doi.org/10.51903/ijgd.v4i1.3699

Tsay, R. S. (2010). *Analysis of financial time series* (3rd ed.). Wiley.

Vaswani, A., Shazeer, N., Parmar, N., Uszkoreit, J., Jones, L., Gomez, A. N., Kaiser, Ł., & Polosukhin, I. (2017). Attention is all you need. In *Advances in Neural Information Processing Systems, 30*, 5998–6008. https://proceedings.neurips.cc/paper/7181-attention-is-all-you-need

Xin, Q. (2025). Uncertainty-aware late fusion for 3D perception (confidence calibration + fusion rule learning). *Journal of Technology Informatics and Engineering, 4*(1), 215–238. https://doi.org/10.51903/jtie.v4i1.485

Xin, Q. (2026a). Explainable and fair credit risk scoring with counterfactual explanations: A reproducible evaluation on the German Credit Dataset (HELOC-motivated). *Journal of Information and Technology, 14*(2), 215–231. https://doi.org/10.32664/j-intech.v14i02.2228

Xin, Q. (2026b). Probabilistic bike-sharing demand forecasting under changing weather and seasonal regimes with transformer-based models. *Findings*. https://doi.org/10.32866/001c.157499

Ye, T., Chang, X., & Zhong, E. (2025). Uncertainty-aware breast ultrasound explanation cards: A visual communication framework for image-based AI diagnostic support using BreastMNIST_224. *International Journal of Graphic Design, 3*(2), 365–380. https://doi.org/10.51903/ijgd.v3i2.3701

Zhang, B., Ren, Y., & Zou, J. (2025). LLM-style explainable e-commerce recommendation cards: A UI/UX design framework for trust-calibrated product recommendation. *International Journal of Graphic Design, 3*(2), 381–396. https://doi.org/10.51903/ijgd.v3i2.3697

Zhang, B., Sun, X., Liu, G., & Zhou, B. (2026). LLM-style DevOps copilot for cloud-native troubleshooting: Retrieval-augmented runbook generation and command-safety evaluation. *Journal of Technology Informatics and Engineering, 5*(2), 104–118. https://doi.org/10.51903/jtie.v5i2.534

Zhang, K., Chen, Y., & Qian, A. (2025). Evidence-grounded accounting disclosure review cards: A visual communication framework for LLM-style explanations over SEC financial statements and notes. *International Journal of Graphic Design, 3*(2), 395. https://doi.org/10.51903/ijgd.v3i2.3710

Zhao, S., Bai, J., & Roberson, D. (2025). Multi-horizon GPU demand forecasting with workload semantics and operational risk curves: An empirical study on Alibaba Clusterdata GPU Trace. *Journal of Technology Informatics and Engineering, 4*(3), 544–571. https://doi.org/10.51903/jtie.v4i3.498

Zhao, S., Ren, Y., & Chang, X. (2026). Profit-aware spot GPU admission control with cost-sensitive loss and evidence-grounded policy memos for AI workload supply-demand matching. *Journal of Technology Informatics and Engineering, 5*(2), 45–59. https://doi.org/10.51903/jtie.v5i2.545

Zhong, Z. S., Chen, J., Zhong, E., & Sun, X. (2025). Evidence-calibrated RAG for unanswerable question answering: Retrieval coverage, abstention calibration, and hallucination-proxy analysis on SQuAD 2.0. *Journal of Technology Informatics and Engineering, 4*(2), 502–520. https://doi.org/10.51903/jtie.v4i2.536

Zhong, Z. S., Wu, Q., & Mi, G. (2025). Uncertainty-aware medical image explanation cards: LLM-generated visual explanations for AI-assisted radiology interfaces. *International Journal of Graphic Design, 3*(2), 415–436. https://doi.org/10.51903/ijgd.v3i2.3616

Zhou, H., & Zhang, K. (2025). News-based uncertainty and macro-market fusion for VIX direction forecasting: Evidence from 2015–2024

FRED panel. *Journal of Technology Informatics and Engineering, 4*(2), 487–501. https://doi.org/10.51903/jtie.v4i2.540

Zhou, S., Chen, Y., & Lee, K. (2026). Accounting-aware evidence-constrained agents for disclosure, settlement, and secondary-market risk monitoring in tokenized RWA infrastructure. *Journal of Technology Informatics and Engineering, 5*(2), 60–74. https://doi.org/10.51903/jtie.v5i2.544