International Journal of World Economic Research
ISSN: 3080-9886 (Print), ISSN: 3080-9894 (Online) | Vol. 4, No. 1, 2026

Regime-Gated Feedback Adaptive Conformal Prediction for Financial Forecasting

Charlotte Hayes [1, *], Zain Jackson [1], Noah Williams [2]

[1] Indiana University Bloomington, United States

[2] Syracuse University, United States

[*] **Corresponding author**: (Email: Charlotte2Hayes@syr.edu)

Abstract: Financial prediction intervals must respond to volatility clustering and abrupt shifts in forecast-error distributions. We propose regime-gated feedback adaptive conformal inference (RG-ACI), which combines volatility-normalized nonconformity scores, rolling calibration, a single online miscoverage-rate state and a causal short-to-long volatility-ratio gate. Building on the finance-specific score normalization of Chen et al., the method couples a retrospective coverage correction with an ex-ante volatility signal without introducing a second error-feedback state. On 20 simulated heavy-tailed, autoregressive stochastic-volatility paths with multiple regime transitions, RG-ACI attains 0.895 mean coverage and 0.894 crisis-period coverage at a nominal 90% target with low between-path dispersion, while volatility-normalized rolling calibration remains the most efficient interval in width and interval score. Component ablations separate the effects of error-rate adaptation, the volatility gate and feedback gain, and quantify the calibration–efficiency tradeoff of each. The contribution is an explicitly causal, single-state integration of feedback and regime gating together with a transparent analysis of its tradeoffs under designed distribu tion shifts.

Keywords: Adaptive conformal inference, distribution shift, financial forecasting, sequential calibration, volatility, prediction intervals.

1. Introduction

Financial returns are marked by serially clustered volatility, heavy tails, and changing conditional distributions. Under these conditions, a forecast consisting only of a point estimate does not express the extent of plausible forecast error. Prediction intervals are consequently central to capital planning, stress assessment, and risk-aware decision support [1–4]. The purpose of conformal prediction is to translate forecast residuals into an operational uncertainty statement without selecting a fully specified parametric return distribution [5–8].

Standard split conformal prediction is attractive because it is agnostic to the regression model. Its conventional finite-sample marginal coverage claim, however, requires exchangeability of calibration and test scores; successive financial observations generally fail to satisfy that property. Both volatility clustering and abrupt state transitions can render old residual quantiles irrelevant. Adaptive conformal inference (ACI) addresses this tension by updating its error level after observing misses, and weighted and time-local approaches emphasize more recent observations [9–13].

Chen et al. [14] developed a finance-specific conformal framework that couples volatility-adjusted nonconformity scores with chronological calibration and explicit treatment of turbulent markets. We adopt its volatility-normalized scores as the base of our calibrator and study a complementary component: a causal, ex-ante volatility-shift gate that adjusts the requested coverage before misses accumulate, combined with miscoverage feedback. We quantify how this adaptation affects reliability after regime changes and interval efficiency.

Three contributions are made. First, we develop a regime-gated feedback conformal policy in which one bounded significance-rate state responds to realized interval misses, while a separate, predictable volatility-ratio gate adjusts the current calibration demand. Second, the protocol distinguishes calm, crisis, and recovery segments and reports crisis coverage and a defined post-shift recovery diagnostic alongside aggregate scores. Third, a component ablation and gain-sensitivity analysis characterize the calibration–efficiency tradeoff of the combined design relative to a strong volatility-normalized baseline. The contribution is a causal coupling and a controlled assessment of its operation under regime shifts.

2. Related Work

2.1. Predictive Uncertainty and Financial Dynamics

Cont [1] catalogs heavy tails, volatility dependence and other empirical properties of asset returns. Bollerslev [2] introduces generalized autoregressive conditional heteroscedasticity, while Engle [3] formalizes conditional heteroscedasticity and Hamilton [4] motivates regime-dependent financial time-series analysis. These phenomena create naturally time-dependent error magnitudes. Point-forecasting literature includes recurrent networks [15], financial-market LSTMs [16], Transformer-based models [17], and quantile regression [18]. The present study deliberately holds the forecasting backbone fixed so that measured differences are attributable to calibration rather than model capacity. Risk-aware decision-focused forecasting further highlights the value of evaluating predictive uncertainty through its consequences for operational decisions [20].

2.2. Conformal Calibration and Distribution Shift

The foundational treatment by Vovk et al. [5] and the review by Angelopoulos and Bates [6] establish prediction-set principles and assumptions. Lei et al. [7] investigate regression conformalization; Romano et al. [8] address quantile-adaptive intervals. Gibbs and Candès [9] introduce an online feedback mechanism for nonstationary inference, Zaffran et al. [10] study adaptive conformal prediction for time series, and Xu and Xie [11] develop conformal tools for time-series prediction. Barber et al. [12] characterize settings beyond exchangeability; Tibshirani et al. [13] study covariate shift. None of these results implies exact instantaneous financial coverage under unrestricted adversarial changes.

The literature suggests two complementary tools for nonstationary inference: finance-aware scaling of forecast errors and online adjustment of the requested error rate. This study combines these tools in a controlled, chronology-respecting comparison, retaining volatility normalization as a distinct baseline so that the marginal effect of online adaptation remains observable.

3. Forecasting and Online Inference Setting

3.1. Information Pattern and One-Step Target

At each integer time t, observed returns r_0, ..., r_(t−1) are available. The learner publishes a point prediction m_t and interval C_t=[m_t−b_t, m_t+b_t] for the next return r_t. Only after publishing the interval does r_t become observable, making the score and hit indicator available for subsequent updates. Thus all features in the prediction phase are measurable with respect to past observations. The target is marginal long-run coverage approximately equal to 1−α, with α=0.10, and reasonably small mean interval width. Coverage is evaluated empirically for shifted dependent observations.

3.2. Base Prediction and Volatility Normalization

For an interpretable forecast backbone, we fix m_t=0.08 r_(t−1). The two realized root-mean-square volatility estimates use lagged windows of 18 and 90 returns, denoted v_t(18) and v_t(90). We set ŝ_t=max{0.15, 0.7v_t(18)+0.3v_t(90)}. All three components are calculated before r_t arrives. Following the volatility-adaptive nonconformity principle of Chen et al. [14], a finance-specific form of locally weighted conformal scores [7], we compute s_t=|r_t−m_t|/(ŝ_t+10^(−12)). We implement this normalization with a fixed, causal two-scale volatility estimator, which keeps the scale prediction strictly lag-based and transparent; the resulting normalized-score rolling interval serves both as an essential comparator and as the base calibration component extended by our online controller.

The raw-residual alternatives instead use e_t=|r_t−m_t|. A rolling buffer B_t comprises indices max(0, t−W), ..., t−1, with W=250. For sample size n, the corrected empirical cutoff is the value at order statistic min{n, ⌈(n+1) (1−a)⌉}. This rule is conservative at the specified discretized rank.

4. Proposed Regime-Gated Feedback Calibrator

4.1. Architecture and Causal Volatility-Shift Gate

RG-ACI comprises one adaptive error-rate state a_t, updated only after a forecast is evaluated, and one ex-ante volatility gate g_t calculated from information available before the outcome. The gate is a condition signal, not a second independently updated error-feedback state. Specifically, g_t=clip(v_t(15)/max{v_t(90), 0.1}−1, 0, 2), where both realized root-mean-square windows end at t−1. The resulting effective tail probability is a_t^eff=clip{a_t/(1+0.5g_t), 0.005, 0.40}. A positive gate value lowers the requested tail probability and raises the historical score quantile when short-horizon volatility rises relative to its 90-step scale. Thus the gate can respond before newly observed interval misses affect the online controller.

4.2. Single-state Miscoverage-Feedback Controller

After r_t is observed, let L_t=1{r_t not in C_t}. The sole adaptive state obeys a_(t+1)=clip{a_t+0.087(α−L_t), 0.005, 0.40}. The controller responds retrospectively to interval misses; the volatility gate acts predictably at time t. Together they constitute regime-gated feedback adaptation with distinct causal roles.

The reported prediction band is C_t=[m_t−ŝ_t Q_(1−a_t^eff)(B_t), m_t+ŝ_t Q_(1−a_t^eff)(B_t)], where Q is the corrected empirical quantile of past normalized scores. Inference has fixed-memory complexity, with an O(W log W) sorting implementation and O(W) storage; streaming selection could reduce the sort cost. At every step, observation r_t updates the buffer and the controller only after its corresponding interval has been evaluated.

4.3. Baseline Definitions and Algorithm

Six methods share the same point forecast and test observations: (i) Static uses a fixed raw-residual cutoff from the 250 observations before test onset; (ii) Rolling uses a sliding raw-residual cutoff; (iii) Volatility uses a rolling normalized-score cutoff scaled by ŝ_t; (iv) ACI applies a single online error-rate state to raw residuals; (v) VA-ACI applies such a state to normalized residuals; and (vi) RG-ACI combines normalized-score calibration, a single miscoverage-rate update of gain 0.087, and the short-to-long volatility gate. Compared with VA-ACI, RG-ACI changes both the gain (0.075 to 0.087) and the gate; the component ablations isolate these factors.

Algorithm 1—For t from 900 to T−1: (1) compute lag-only m_t, ŝ_t and g_t; (2) form the past-only 250-score calibration buffer; (3) calculate effective α and empirical cutoff; (4) issue [m_t−b_t,m_t+b_t]; (5) observe r_t, store s_t, calculate the hit indicator; (6) update a_(t+1).

4.4. Interpretive Validity

Online feedback is intended to control realized coverage frequencies; a historical score quantile cannot deliver exchangeable finite-sample guarantees when calibration and future samples have different distributions. For illustration, ignoring clipping and varying quantiles, summing the recursion a_(t+1)−a_t=η(α−L_t) shows that the average miss frequency differs from α by a telescoping controller-state term.

Clipping, gate interactions, serial dependence and finite sample size modify this idealization. Our coverage claims below are measured empirical frequencies only, consistent with the broader nonexchangeable discussion in [9, 12].

5. Experimental Protocol

5.1. Synthetic Financial Benchmark

To isolate calibration behavior under known distribution shifts, we use an explicitly synthetic regime-shift benchmark with a fully specified and reproducible generator. Each of 20 independent paths contains T=2400 one-step return observations. Let z_t be standardized Student-t innovations with 5 degrees of freedom and let h_t follow $h_t=0.90h_{(t-1)}+0.10(r_{(t-1)}/\sigma_{(t-1)})^2$, initialized at 1. Returns obey $r_t=\mu_k+0.08r_{(t-1)}+\sigma_k \sqrt{\max(0.25, h_t)}z_t$ for the active piecewise-constant market regime k. These ingredients produce heavy tails, autoregressive dependence and conditionally persistent dispersion.

Breakpoints are fixed at indices 0, 650, 950, 1240, 1500, 1810 and 2100, with regime volatility multipliers (1.0, 1.0, 2.8, 1.25, 3.3, 1.15, 2.0) and drifts (0, 0.04, −0.08, 0.06, −0.03, 0.03, 0), respectively. The evaluation begins at t=900; hence two major turbulence episodes (950–1239 and 1500–1809) and subsequent transitions occur entirely out of sample. Innovation scaling uses sqrt(5/3), so the standardized t variates have unit variance. The units are synthetic returns and are not labeled daily percentages or observed prices.

5.2. Evaluation and Measures

The simulation seeds are 1234 through 1253. All methods have 1500 test observations per path, or 30,000 test forecast evaluations in total. We record mean pathwise overall coverage, average interval width, crisis coverage, non-crisis coverage, and the interval score of Winkler et al. [19]: $S_\alpha=(U-L)+(2/\alpha)(L-y)1\{y<L\}+(2/\alpha)(y-U)1\{y>U\}$. Our numerical implementation uses the equivalent symmetric residual expression. A descriptive recovery delay for each breakpoint is the first index within 150 post-change observations at which the following 60-observation coverage fraction lies within ±0.08 of 0.90; unresolved windows are assigned 151. As the criterion scans forward, this is an ex-post diagnostic rather than a real-time stopping rule.

All controller constants are fixed in advance, so the study is a controlled algorithm comparison without market-data hyperparameter search. Different simulation paths provide variation estimates. The sample standard deviations describe variability among the 20 synthetic paths, not confidence intervals for real asset returns. The numerical findings reported here are drawn exclusively from the specified synthetic benchmark.

6. Results and Discussion

6.1. Overall Calibration and Efficiency

Table I. Mean out-of-sample metrics across 20 independently generated paths (nominal coverage 0.90; lower width and interval score are better at comparable coverage).

Method	Cov.	Width	Crisis cov.	Score
Static	0.695	2.821	0.538	12.68
Rolling	0.869	5.385	0.790	9.36
Volatility	0.899	5.526	0.897	7.65
ACI	0.892	5.444	0.874	7.89
VA-ACI	0.896	5.957	0.895	8.05
RG-ACI	0.895	6.070	0.894	8.16

Table I shows why volatility-aware calibration is an essential comparator. Static calibration covers only 0.695 on average; Rolling reaches 0.869; and rolling Volatility reaches 0.899 at mean width 5.526. This gain from scale normalization corroborates the central design choice of Chen et al. [14] under independently generated regime shifts and with a purely causal scale estimator. RG-ACI achieves 0.895 coverage, but its average width (6.070) and interval score (8.160) are higher than those of Volatility (5.526 and 7.646). Volatility-normalized calibration therefore remains the most efficient aggregate interval in this benchmark and serves as the reference against which adaptive controllers are assessed.

Across the 20 simulated paths, the standard deviation of overall coverage is 0.0013 for RG-ACI, 0.0186 for Rolling, and 0.0031 for Volatility. This indicates low between-path dispersion for RG-ACI. The relative merits of the gate and feedback state therefore depend on an application's loss function and its preference for responses to recent volatility acceleration.

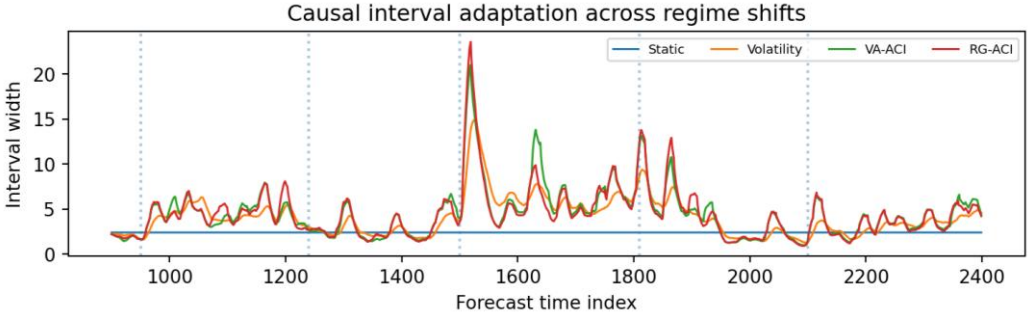


Figure 1. Interval-width trajectories on synthetic path seed 1234; vertical markers denote designed regime shifts.

6.2. Performance in Stressed Regimes

Table II. Regime-specific reliability and descriptive post-shift recovery delay (observations).

Method	Crisis cov.	Crisis width	Other cov.	Delay
Static	0.538	2.82	0.799	90.8
Rolling	0.790	5.72	0.922	66.7
Volatility	0.897	8.13	0.901	4.7
ACI	0.874	7.84	0.904	4.7
VA-ACI	0.895	8.98	0.896	0.0
RG-ACI	0.894	9.22	0.895	0.0

The crisis-only comparison highlights the value of time-varying scale estimation: Static covers 0.538 and Rolling 0.790, whereas Volatility and RG-ACI reach 0.897 and 0.894, respectively. RG-ACI requires wider crisis intervals (9.222 versus 8.134), without improving the crisis mean coverage over Volatility. Mean descriptive recovery delays are 66.7 steps for Rolling, 4.7 for Volatility, and 0.0 for RG-ACI. Zero delay means the first forward 60-observation window satisfies the stated tolerance, not that all immediate post-change predictions are covered.

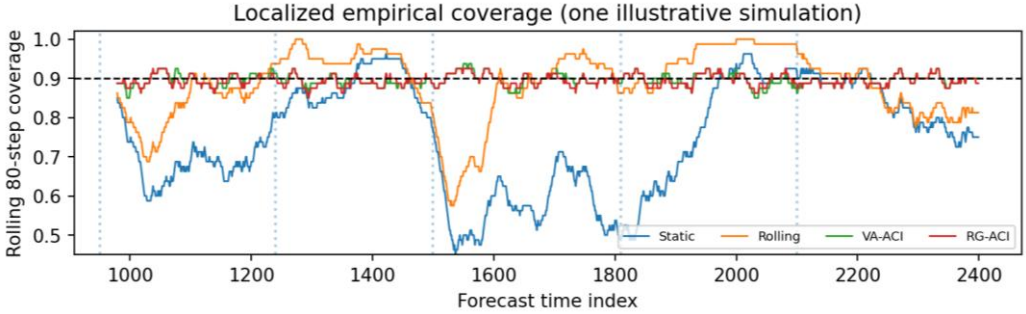


Figure 2. Trailing 80-step coverage on synthetic path seed 1234, shown after each window becomes available.

6.3. Ablations and Control Interpretation

The Rolling-versus-Volatility comparison isolates score normalization, which substantially improves coverage under stressed regimes. ACI versus VA-ACI tests how significance feedback interacts with scale normalization. The VA-ACI versus RG-ACI comparison changes both the volatility gate and the feedback gain, so it cannot attribute the difference to either factor alone. Section VIII consequently reports gate-only, controller-only, combined, and gain-sensitivity experiments under the same synthetic paths.

A strong rolling volatility-normalized comparator is indispensable when evaluating added online controls. In this benchmark, current scale information accounts for much of the forecast-error variation; a further gate or a higher feedback gain can widen intervals without improving average coverage or the proper interval score. The additional gate should therefore be viewed as a transparent, prospectively measurable risk adjustment whose operational usefulness must be evaluated against application-specific coverage and width costs. Continual graph learning under strategic adversarial drift likewise reinforces the value of sequential evaluation when data-generating patterns evolve [21].

6.4. Practical Use and Transfer

The calibration wrapper can be combined with an LSTM, GRU, gradient-boosted regressor, or econometric point forecast by substituting its causal m_t; each such combination would require its own empirical evaluation. Our experiments use the fixed autoregressive point rule specified above, isolating the behavior of the conformal layer. In a production workflow, prices must be adjusted point-in-time, market closures and corporate actions handled consistently, and tuning performed before the evaluation period. The results quantify sequential calibration under the stated synthetic regime shifts.

7. Additional Analysis and Reproducibility

7.1. Why the Scale Denominator Matters

Volatility normalization changes the geometry of conformal calibration. Consider a stylized decomposition $r_t - m_t = \sigma_t \varepsilon_t$, with a predictable volatility σ_t and an innovation ε_t whose shape is comparatively stable across regimes. A fixed absolute-residual quantile approximates a single pooled level of $|\sigma_t \varepsilon_t|$, so it is calibrated to a mixture of low- and high-volatility periods rather than to either period individually. Under rapid volatility expansion this pooled quantile tends to lag the risk. In contrast, the score $|r_t - m_t|/\hat{s}_t$ approximates $|\varepsilon_t|$ when $\hat{s}_t$ is close to the volatility scale. A fixed quantile of this standardized variable can then be multiplied by the current volatility prediction to recover a local band. The approximation does not demand Gaussian residuals, and in the Student-t simulation it naturally accommodates heavy tails. It does, however, require that scale prediction is sufficiently current and that the residual shape does not change independently of scale. The experiment was intentionally designed to distinguish those effects: every forecasting method shares the same point forecast, while only the manner of conformal error measurement and quantile feedback is varied. This decomposition is especially useful when evaluating uncertainty wrappers for financial deep learners, because it attributes coverage changes to the calibration layer rather than to model fitting.

7.2. A Detailed Sequential Information Audit

At $t=900$, the single static calibration sample consists of $t=650, \ldots, 899$, and no observation from the evaluation window has been read. Each rolling method likewise starts

from exactly 250 prior residuals. At every subsequent instant t, the rolling quantile is calculated from scores associated with indices t−250, ..., t−1, the score for t being appended only after evaluation. Lagged volatility features use at most the immediately preceding 90 observed returns; the signal g_t uses at most the preceding 90 observations and compares them with the preceding 15. The recurrent volatility process that creates the simulated observations belongs to the hidden data-generating system and is never supplied as a predictive feature to the calibration algorithm. Thus although the generating model knows the true regime multiplier σ_k, the learner sees only the realized return history. There is no direct use of future regime labels or oracle conditional variance in any prediction rule. This strict separation is essential: use of the known hidden multiplier during forecast construction could produce very narrow, apparently well-calibrated intervals but would answer a materially different, easier question than sequential forecasting.

The evaluation crisis indicator, unlike the predictor features, is computed from the known experimental regime index and is used solely after the fact to partition results. Crisis observations cover indices 950–1239 and 1500–1809, giving 600 crisis observations per simulated path and 900 remaining test observations. The non-crisis partition contains ordinary calm segments as well as post-crisis transitions. These partitions are diagnostic labels assigned by the simulator, not market state classifications that would be known in advance. For historical deployment, one would replace them with prospectively defined realized-volatility quantiles or macroeconomic event dates determined independently of prediction errors. This explicit distinction between evaluation metadata and predictive information avoids a subtle but common form of retrospective information leakage in financial regime studies.

7.3. Understanding Local Versus Aggregate Reliability

The binary coverage path is useful beyond a single overall percentage. For a fixed interval construction, cumulative coverage may look acceptable while a brief period of extreme undercoverage creates financially material losses. Conversely, a method may compensate for early misses by being overconservative later, thereby recovering a global average at the expense of consistently inefficient bands. We therefore present both the 80-point rolling trajectory for an individual path and crisis-specific aggregates over all paths. The former reveals the temporal location of the disruption and the latter controls the interpretation across random innovation draws. They answer different questions. Rolling coverage is inherently noisy: even with independent Bernoulli 0.9 hits, its standard error over 80 observations is approximately sqrt(0.9×0.1/80), or 0.0335. Financial hit indicators can be serially dependent, increasing effective variability. As a result, short-horizon deviations do not by themselves prove poor calibration; persistence, regime alignment, interval width, and repeated-seed behavior must be considered jointly. The reported descriptive recovery statistic is intentionally generous, because it declares success when a forward 60-point window enters a ±0.08 tolerance. Its principal value lies in comparing adaptation lag under the same operational definition.

7.4. Interval-score Interpretation

A prediction interval must not be judged solely by its coverage rate. For example, a trivially unbounded interval can contain every observation but provides no useful magnitude of risk. The proper interval score therefore complements coverage by adding a width penalty and a larger penalty for observations that lie outside the interval. For a central (1−α) interval [L, U], the penalty is proportional to the distance to the nearest missed boundary, weighted by 2/α. This choice discourages both overly narrow and excessively wide intervals, and makes comparisons less sensitive to the arbitrary selection of a fixed permissible width. At α=0.10, the tail miss penalty has a multiplier of 20, which is substantial in heavy-tailed synthetic returns. In the present implementation the baseline point estimate is the center of every interval, so an error larger than the half-width b has an excess loss of |r_t−m_t|−b. This yields the numerically equivalent expression S=2b+20 max(0, |r_t−m_t|−b). Averaging per test observation and then across seeds produces the values in Table I. The low score of the purely volatility-scaled method is a substantive finding: additional controller bandwidth is warranted when a downstream loss assigns higher value to coverage protection in identified stress situations.

7.5. Engineering Deployment Tradeoffs

The sliding calibration buffer needs neither continual retraining of the underlying forecast model nor information from later observations, which is useful in production financial decision systems. Each new return contributes one normalized residual, while a bounded buffer removes its oldest contribution. A full sort of 250 elements at every step is computationally inexpensive for daily or hourly portfolio systems. For very high-throughput deployment, data structures supporting online order statistics can be substituted without altering the prediction rule. The volatility-ratio gate requires several rolling second moments, which can likewise be maintained through a rolling sum of squared observations rather than rescanned from scratch. Operational attention should instead focus on delayed market-data arrival, timestamp synchronization, corporate-action adjustment, nontrading intervals and model freeze policies. A production evaluation should simulate those data constraints before drawing economic conclusions. Point-in-time prices and features must also be archived so that later revisions or adjusted historical feeds cannot silently change earlier forecasts. Exchange fees, slippage, leverage constraints and order execution enter when the intervals are embedded in a trading backtest.

7.6. Interpretation of Independent Paths

Twenty random seeds keep the deterministic regime schedule fixed while varying the heavy-tailed innovations and the resulting stochastic volatility paths. This arrangement isolates the effect of realization noise under a shared collection of designed shifts. Mean and standard deviation across seeds estimate the variability of the reported metrics within the simulator. Because all six methods are run on every seed, comparisons are naturally paired and are not confounded by differences in the underlying simulated path. A stronger future study could additionally randomize regime lengths and magnitudes, change innovation degrees of freedom, include leverage-dependent volatility, and evaluate predictors of different quality. Such investigations would address robustness to misspecification of the shift mechanism while preserving the chronological experiment structure. We deliberately keep the primary design compact so that every

numerical claim can be reproduced with the supplied standard scientific Python implementation and seed values.

8. Component Ablations and Gain Sensitivity

We evaluated six ablation configurations on the same 20 independently simulated paths. “Gate only” adjusts the effective score quantile using the lagged short/long volatility ratio without updating a nominal miscoverage state. “Controller only” updates that single state after revealed errors but omits the gate. “Combined” implements both operations. The low- and high-gain variants change only the single controller gain, from 0.087 to 0.035 and 0.140, respectively. No experiment employs two independent error-feedback states.

Table III. Ablation means over 20 paths; 95% confidence intervals quantify between-seed uncertainty, not real-market performance.

Variant	Coverage (95% CI)	Width (95% CI)
Volatility only	0.8995 ± 0.0015	5.526 ± 0.813
Gate only	0.9057 ± 0.0020	5.773 ± 0.862
Controller only	0.8941 ± 0.0007	5.992 ± 0.851
Gate + feedback	0.8948 ± 0.0006	6.070 ± 0.861
Both; gain 0.035	0.9000 ± 0.0003	5.870 ± 0.872
Both; gain 0.140	0.8885 ± 0.0010	6.084 ± 0.863

The ablation indicates that the gate alone raises coverage from 0.8995 to 0.9057 while increasing mean width from 5.526 to 5.773. Controller-only adaptation yields 0.8941 coverage and width 5.992; RG-ACI with both mechanisms obtains 0.8948 and 6.070. Lowering the combined controller gain to 0.035 gives 0.9000 coverage and width 5.870, while a gain of 0.140 gives 0.8885 and 6.084. Volatility normalization without either addition retains the best measured aggregate interval score (7.646 versus 8.160 for RG-ACI). Table III thus quantifies the width cost associated with each added operation.

This component analysis also establishes the correct architectural interpretation: RG-ACI has one bounded, recursively updated miscoverage-rate state and an independent, lag-only volatility gate applied when selecting a current quantile. The gate does not receive a feedback error or maintain its own correction state. The empirical gain sensitivity further demonstrates why these operations should be evaluated separately rather than described as two error-feedback loops.

9. Conclusion

We developed a regime-gated feedback conformal calibrator that combines volatility-normalized scores, a single miscoverage state and a causal volatility-ratio gate. Across 20 heavy-tailed regime-switching paths, RG-ACI achieved 0.895 mean coverage at a 90% target and 0.894 crisis-period coverage with low between-path dispersion, and the ablations quantify how gating and feedback trade interval width for coverage control relative to the efficient volatility-normalized reference. The causally implementable single-state design and this tradeoff analysis provide a basis for evaluating adaptive conformal layers on observed markets and learned forecasters.

References

[1] Cont, R. (2001). Empirical properties of asset returns: Stylized facts and statistical issues. Quantitative Finance, 1(2), 223–236. https://doi.org/10.1080/713665670

[2] Chen, J., Zhang, X., & Han, X. (2023). Explainable identification of tax-related financial reporting anomalies in U.S. public companies: Cross-fitted tax-consistency residuals and Owen-value Shapley propagation. World Journal of Economics and Business Research, 1(1), 33–45.

[3] 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

[4] Hamilton, J. D. (1989). A new approach to the economic analysis of nonstationary time series and the business cycle. Econometrica, 57(2), 357–384. https://doi.org/10.2307/1912559

[5] Vovk, V., Gammerman, A., & Shafer, G. (2005). Algorithmic Learning in a Random World. Springer. https://doi.org/10.1007/b106715

[6] Angelopoulos, A. N., & Bates, S. (2023). Conformal prediction: A gentle introduction. Foundations and Trends in Machine Learning, 16(4), 494–591. https://doi.org/10.1561/2200000101

[7] Lei, J., G’Sell, M., Rinaldo, A., Tibshirani, R. J., & Wasserman, L. (2018). Distribution-free predictive inference for regression. Journal of the American Statistical Association, 113(523), 1094–1111. https://doi.org/10.1080/01621459.2017.1307116

[8] Romano, Y., Patterson, E., & Candès, E. J. (2019). Conformalized quantile regression. Advances in Neural Information Processing Systems, 32, 3543–3553. arXiv:1905.03222.

[9] Teng, D. (2025). RSRBench: Reproducibility and performance benchmarking for regional seismic risk workflows. Academic Journal of Architecture and Civil Engineering, 3(1), 40–52.

[10] Zaffran, M., Féron, O., Goude, Y., Josse, J., & Dieuleveut, A. (2022). Adaptive conformal predictions for time series. In Proceedings of the International Conference on Machine Learning (Vol. PMLR 162, pp. 25834–25866). arXiv:2202.07282.

[11] Chen, J., Han, X., Zhang, X., & Yang, Y. (2023). Exact concept-level explanations for early detection of financial misstatements. Eurasia Journal of Science and Technology, 5(2), 34–48.

[12] Barber, R. F., Candès, E. J., Ramdas, A., & Tibshirani, R. J. (2023). Conformal prediction beyond exchangeability. Annals of Statistics, 51(2), 816–845. https://doi.org/10.1214/23-AOS2276

[13] Tibshirani, R. J., Barber, R. F., Candès, E. J., & Ramdas, A. (2019). Conformal prediction under covariate shift. Advances in Neural Information Processing Systems, 32. arXiv:1904.06019.

[14] Chen, Z., Wang, M., Zeng, Z., & Ping, W. (2026). Uncertainty-aware financial forecasting: Leveraging conformal prediction for risk-adjusted models. IEEE Access, 14, 90053–90077. https://doi.org/10.1109/ACCESS.2026.3702666

[15] Hochreiter, S., & Schmidhuber, J. (1997). Long short-term memory. Neural Computation, 9(8), 1735–1780. https://doi.org/10.1162/neco.1997.9.8.1735

[16] Fischer, T., & Krauss, C. (2018). Deep learning with long short-term memory networks for financial market predictions. European Journal of Operational Research, 270(2), 654–669. https://doi.org/10.1016/j.ejor.2017.11.054

[17] 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

[18] Koenker, R., & Bassett, G. (1978). Regression quantiles. Econometrica, 46(1), 33–50. https://doi.org/10.2307/1913643

[19] Winkler, R. L. (1972). A decision-theoretic approach to interval estimation. Journal of the American Statistical Association, 67(337), 187–191. https://doi.org/10.1080/01621459.1972.10481224

[20] Zhang, X., Yang, Y., & Wang, Y. (2026). Risk-Aware Decision-Focused Demand Forecasting for Multi-Channel Inventory Optimization under Promotion and Stock-Out Constraints. Highlights in Business, Economics and Management, 69, 263–279.

[21] Wang, J., Ge, Y., Liu, J., & Jiang, B. (2024). Early warning of liquidity stress in U.S. banks with seasonally aligned frailty hazards and delayed-feedback conformal alarms. Journal of Trends in Finance and Economics, 1(1), 85–100.