Journal of Mathematical Finance and Risk Management
ISSN: 3080-9983 (Print), ISSN: 3080-9991 (Online) | Vol. 1, No. 2, 2026

Regime-Aware Conformal Transformer for Multi-Horizon Financial Forecasting Under Market Uncertainty

Julia Stein

Texas A&M University, Department of Computer Science and Engineering, USA
Julia.steinedu@outlook.com

Abstract: Financial forecasting systems deployed in portfolio monitoring and risk control require not only point forecasts, but also prediction intervals that remain meaningful when markets change volatility regimes. This paper proposes a Regime-Aware Conformal Transformer (RACT) for multi-horizon financial forecasting under market uncertainty. The model combines a compact Transformer encoder with asset and volatility-regime tokens, producing simultaneous return forecasts for 1-, 2-, 4-, and 8-week horizons. To quantify uncertainty, we introduce a regime-aware conformal calibration rule that computes horizon-specific volatility-normalized residual quantiles inside low-, medium-, and high-volatility regimes, while retaining a global conformal safety floor to avoid under-coverage when local volatility scaling becomes too optimistic. The resulting interval construction is conservative, causally ordered, and easy to reproduce. Experiments use the real Plotly Express stocks dataset containing weekly normalized closing prices for six technology stocks in 2018/2019. On this compact public dataset, RACT with regime-aware conformal calibration achieves 98.3% average empirical coverage at a 90% nominal level and improves 8-week coverage from 76.5% for pure volatility-scaled conformal calibration to 94.6%, at the cost of wider intervals. The results show that regime-aware conformal envelopes can materially improve long-horizon reliability, although they should be validated on larger trading datasets before production deployment.

Keywords: Conformal prediction; Financial forecasting; Market regimes; Multi-horizon forecasting; Transformer; Uncertainty quantification.

1. Introduction

Financial return forecasting is difficult because the object being predicted is not a stable physical process but a market-clearing outcome shaped by changing liquidity, macroeconomic news, investor positioning, and feedback between models and trading decisions. A useful forecasting model therefore has to answer two questions at the same time: what is the expected direction or magnitude of the next movement, and how wide is the range of plausible outcomes? Classical point forecasts are incomplete for this setting because an apparently accurate prediction can become dangerous when its uncertainty is underestimated. This problem is central to modern risk-aware forecasting, where decisions are frequently constrained by downside risk rather than by mean prediction alone.

Recent uncertainty-aware financial forecasting research has shown that conformal prediction can provide a practical bridge between machine-learning predictors and statistically calibrated prediction intervals. In particular, Chen et al. developed a volatility-adaptive conformal framework for financial forecasting and demonstrated the value of combining temporal calibration, volatility adaptation, and extreme-event awareness for risk-adjusted forecasts [1]. Their work is an important foundation for this paper because it frames uncertainty quantification as a first-class forecasting objective rather than as a post-hoc visualization. However, their main backbone families are recurrent models and related forecasters, whereas many current multi-horizon forecasting systems increasingly rely on attention-based architectures.

This paper extends the same uncertainty-aware philosophy toward a Transformer-based, regime-aware multi-horizon setting. The proposed Regime-Aware Conformal Transformer, abbreviated RACT, uses a compact Transformer encoder to learn temporal representations from weekly return windows and injects both asset identity and volatility-regime information through learnable tokens. The methodological focus is not to claim that a small public dataset can establish a profitable trading strategy. Instead, the paper studies a reproducible uncertainty-forecasting pipeline that can be executed end-to-end: data loading, causal splitting, model training, conformal calibration, interval evaluation, and figure generation.

The main contribution is a conservative regime-aware conformal calibration method for multi-horizon forecasts. For each horizon, calibration residuals are normalized by local volatility scaled by the square root of the horizon. Residual quantiles are then estimated inside low-, medium-, and high-volatility regimes. Because regime-specific calibration cells can be small or unstable, especially in short public datasets, the proposed interval also includes a global absolute conformal floor. This floor preserves a simple split-conformal safety envelope while allowing the volatility-regime component to widen intervals during turbulent states. The empirical results confirm that this design is especially useful at longer horizons, where pure volatility scaling can underestimate uncertainty under drift.

The paper is designed with three practical constraints in mind. First, the model must be causal: every feature, regime label, and calibration score used at a forecast origin must be available before the future return is observed. Second, the experiment must be auditable: every number in the tables should be regenerated from the provided code rather than entered manually. Third, the contribution should remain useful even when point forecasts are weak. This final constraint is important in finance, because a risk-control model can be operationally valuable if it produces honest intervals and warning ranges, even if it does not produce a profitable directional signal by itself.

The proposed framework should therefore be viewed as an uncertainty layer around a multi-horizon forecasting model. Its purpose is not to replace fundamental analysis or portfolio construction, but to provide calibrated ranges for downstream processes such as position sizing, stop-loss review, scenario screening, and risk committee reporting. In such settings, the interval width is not merely a statistical artifact; it is a communication channel that tells the decision maker when the model is operating in a familiar regime and when it is extrapolating under elevated market uncertainty.

The remainder of the paper follows the structure expected in an engineering conference article. Section II reviews the relevant literature. Section III presents the RACT methodology and conformal interval construction. Section IV describes the reproducible experimental design. Section V reports empirical results and explains the observed trade-offs between coverage and interval width. Section VI discusses limitations and future extensions, and Section VII concludes the study.

2. Related work

Financial forecasting has a long tradition in econometrics and risk management. Markowitz's portfolio selection framework formalized the relationship between expected return and variance [2], while Engle's ARCH model and Bollerslev's GARCH model established conditional heteroskedasticity as a central property of financial time series [3], [4]. Hamilton's regime-switching model further emphasized that economic and financial dynamics can move between latent states with different parameters [5]. These ideas remain relevant because financial returns exhibit heavy tails, volatility clustering, and temporal dependence, which Cont summarized as stylized empirical facts of asset returns [6].

Deep learning methods have expanded the modeling toolkit for financial time series. LSTM networks have been applied to large-scale equity movement prediction [7], and systematic reviews report broad adoption of recurrent, convolutional, and attention-based architectures in price and return forecasting [8]. The Transformer architecture introduced self-attention as an efficient mechanism for sequence modeling [9]. More specialized time-series models such as Temporal Fusion Transformers, PatchTST, TimesNet, and DeepAR have demonstrated that neural forecasting can combine temporal representation learning with multi-horizon or probabilistic outputs [10]–[13]. Nevertheless, accurate point prediction remains extremely challenging in liquid financial markets, and strong naive baselines may perform competitively when returns are close to martingale differences.

Conformal prediction offers a distribution-free way to wrap black-box predictors with prediction sets or intervals. The foundations were developed by Vovk, Gammerman, and Shafer [14], and later tutorials and split-conformal formulations made the approach accessible for practical regression tasks [15], [16]. Conformalized quantile regression combines conditional quantile models with conformal calibration to obtain adaptive intervals under exchangeability [17]. More recent work studies conformal inference beyond exchangeability, adaptive conformal inference under distribution shift, and conformal forecasting for time series [18]–[21]. These studies motivate the design of time-aware calibration rules for financial data, where strict exchangeability is rarely plausible.

This paper is closest to the uncertainty-aware financial forecasting framework of Chen et al. [1], which argues for volatility-adaptive nonconformity measures and temporal conformal calibration in financial forecasting. We build positively on that insight but introduce a different architectural and calibration emphasis: a Transformer encoder with regime tokens, multi-horizon return outputs, and a conservative hybrid conformal rule that combines regime-specific volatility scaling with a global conformal safety floor. The result is a concise and reproducible framework targeted at the practical problem of interval reliability under changing market regimes.

Uncertainty quantification in financial forecasting can be divided into model-internal and model-external approaches. Model-internal methods, such as Bayesian neural networks, Monte Carlo dropout, quantile regression, or deep ensembles, attempt to learn a predictive distribution directly. These approaches are flexible, but their uncertainty estimates may be miscalibrated when the model class is misspecified or when the test period differs from the training period. Model-external methods, such as conformal prediction, instead calibrate residual behavior on a held-out sample and can wrap around many different forecasting backbones. This modularity is attractive for financial systems because institutions often need to change the point model without redesigning the entire risk-control layer.

A second relevant line of work concerns multi-horizon forecasting. Short-horizon forecasts and long-horizon forecasts are not simply versions of the same problem with different labels. Longer horizons aggregate more shocks, are more exposed to regime transitions, and often have different residual distributions. Attention-based models such as TFT and PatchTST address representation learning for multi-step prediction, but their raw outputs still require calibration when the forecast is used in risk-sensitive decisions. RACT addresses this by estimating horizon-specific conformal radii rather than using a single residual scale across all forecast horizons.

Finally, market-regime modeling is a natural complement to conformal prediction. Financial practitioners routinely distinguish between calm and turbulent markets because the same absolute forecast error has different meaning in different volatility states. The proposed method operationalizes this intuition through transparent rolling-volatility regimes. More complex alternatives are possible, but simple volatility terciles have two advantages for a reproducible conference paper: they are easy to audit and they avoid hidden future leakage.

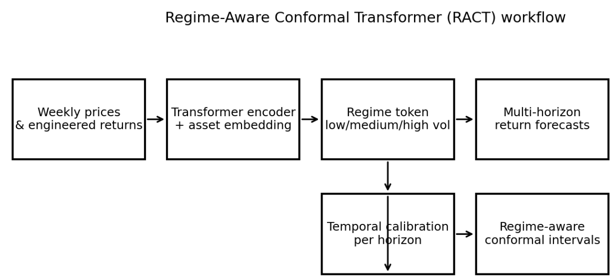

Fig. 1 RACT workflow: regime-aware Transformer forecasting followed by temporal conformal calibration.

3. Methodology

Let $P_{a,t}$ denote the normalized closing price of asset a at weekly time t, and let $r_{a,t}=\log(P_{a,t})-\log(P_{a,t-1})$ be the log return. Given a lookback window of length L, the input $X_{a,t}$ contains lagged returns, short rolling mean, short and medium rolling volatility, four-week momentum, and cross-sectional market return features computed only from information available at or before t. The forecasting target is the vector $y_{a,t}=(r_{a,t+1}, r_{a,t+2}, r_{a,t+4}, r_{a,t+8})$, where each component is a cumulative future log return from t to the corresponding horizon. This formulation avoids using future information in features and makes the multi-horizon nature of the task explicit.

The forecasting backbone is a compact Transformer encoder. Each feature vector in the lookback sequence is projected into a d-dimensional token embedding, positional embeddings are added, and the sequence is processed by a self-attention encoder. Two additional learnable embeddings are included. The asset embedding allows the model to represent persistent cross-asset differences, while the regime embedding informs the model whether the current forecast origin belongs to a low-, medium-, or high-volatility state. Regimes are assigned by training-set terciles of rolling eight-week volatility, so test-time labels are computed without looking ahead. The final hidden state is passed to a multi-output head that predicts all horizons simultaneously.

The conformal layer transforms point forecasts into prediction intervals. For calibration sample i and horizon h, the ordinary residual is $e_{i,h}=|y_{i,h}-f_h(X_i)|$. A global split-conformal radius q^{abs}_h is computed from these residuals. In parallel, a volatility-normalized score is defined as $s_{i,h}=e_{i,h}/(\sigma_i \sqrt{h}+\epsilon)$, where σ_i is the local rolling volatility and ϵ is a small stability constant. For each volatility regime g, a regime-specific quantile $q_{g,h}$ is estimated from the scores in that regime and shrunk toward the global normalized quantile to reduce small-cell noise.

For a new forecast origin in regime g, the final half-width is $w_{g,h}(X)=\max(q^{abs}_h, q_{g,h}\, \sigma(X) \sqrt{h})$. The prediction interval is $C_h(X)=[f_h(X)-w_{g,h}(X), f_h(X)+w_{g,h}(X)]$. This construction has two advantages. First, the global conformal floor prevents the volatility-scaled component from becoming narrower than a standard conformal interval for the same point model. Second, the regime-volatility component can still expand the interval when current volatility indicates elevated uncertainty. Under exchangeability, the global floor inherits the usual split-conformal marginal coverage argument; under financial non-stationarity, it should be interpreted as a conservative practical safeguard rather than a strict guarantee for arbitrary regime shifts.

Algorithmically, the pipeline proceeds in five stages: feature construction, temporal train/validation/calibration/test splitting, Transformer training with validation-based early stopping, conformal quantile estimation on the calibration block, and interval evaluation on the held-out test block. The procedure is intentionally simple because the main objective is reproducible interval calibration rather than an over-engineered trading model.

The regime definition is intentionally minimal. Let $\sigma_{a,t}$ be the rolling eight-week standard deviation of weekly returns for asset a. Two thresholds are estimated from training forecast origins only: the 33rd percentile and the 66th percentile of sigma. Forecast origins below the first threshold are labeled low-volatility, origins between the thresholds are labeled medium-volatility, and origins above the second threshold are labeled high-volatility. The thresholds are then frozen and applied to validation, calibration, and test origins. This design prevents the test distribution from influencing the regime partition.

The asset and regime embeddings serve different purposes. Asset embeddings capture persistent differences between the six series, such as typical volatility level or trend behavior during the sample. Regime embeddings capture the market state at the forecast origin and let the same lag pattern be interpreted differently in a quiet or turbulent environment. Because the model is compact, these embeddings are not expected to learn deep economic structure; their role is to provide a controlled inductive bias that is visible in the architecture and easy to ablate.

The conservative conformal envelope can be interpreted as a union of two interval logics. The global radius protects the method from small-sample errors in regime-specific calibration. The volatility-regime radius protects the method from treating all states as equally uncertain. Taking the maximum of the two radii is simple but effective: it never produces intervals narrower than global CP for the same point model, and it can become wider when the regime-specific volatility score indicates larger uncertainty. This is particularly important at the 8-week horizon, where local volatility scaling alone underestimates risk in the reported experiment.

The computational complexity is dominated by Transformer training. Calibration is inexpensive because it only requires residual computation and empirical quantiles over the calibration block. In the supplied implementation, the entire experiment, including two neural seeds, deterministic baselines, CSV export, and figure generation, runs on CPU. This is important for reproducibility: readers do not need a specialized GPU server to verify the tables.

Algorithm 1 summarizes the procedure in prose. First, the stock panel is transformed into lagged feature tensors and multi-horizon targets. Second, samples are split by forecast-origin date. Third, a Transformer is trained on the training block and selected by validation loss. Fourth, calibration residuals are computed for each horizon. Fifth, global absolute residual quantiles and regime-specific volatility-normalized quantiles are estimated. Sixth, test intervals are formed using the maximum of the global radius and the regime-volatility radius. Finally, coverage, width, and Winkler score are computed without any manual adjustment.

Step	Operation	Leakage control
1	Build causal features and targets	Uses history up to t only
2	Split by forecast-origin date	No random mixing across dates
3	Train Transformer backbone	Validation only for early stopping
4	Compute calibration residuals	Calibration block is after validation
5	Estimate global and regime quantiles	No test labels used
6	Evaluate test intervals	Held-out test block only

Algorithm 1. Reproducible RACT training and calibration pipeline.

4. Experimental design

The experiment uses the real Plotly Express stocks dataset, whose official documentation describes it as closing prices from six technology stocks in 2018/2019 [22]. The six assets are GOOG, AAPL, AMZN, FB, NFLX, and MSFT. The dataset contains 105 weekly observations and is normalized rather than raw dollar-priced, making it suitable for a compact reproducibility experiment but not a comprehensive market backtest. The script included with this paper exports the raw data, engineered samples, train logs, calibration outputs, test predictions, and all result tables used in the manuscript.

A 12-week lookback window is used. Forecast origins are split by date rather than randomly: 50% for training, 15% for validation, 15% for conformal calibration, and 20% for testing. This yields 252 training samples, 78 validation samples, 78 calibration samples, and 102 test samples after stacking the six assets. Features are standardized using training statistics only. The Transformer uses one encoder layer, a 24-dimensional token size, four attention heads, AdamW optimization, and early stopping on validation loss. Neural results are averaged over two fixed random seeds, while deterministic baselines have one run.

The baselines are selected to represent both simple and model-based forecasting. Persistence predicts zero future log return and is a strong baseline for short-horizon equity returns. Ridge regression uses flattened lag features with global conformal intervals. Transformer + global CP uses the same Transformer style without regime tokens. RACT + global CP isolates the effect of adding regime tokens to point forecasts while retaining ordinary global conformal calibration. RACT + volatility-scaled CP uses volatility-normalized scores without the global safety floor. The proposed method is RACT + regime-aware CP.

Evaluation uses root mean squared error (RMSE), mean absolute error (MAE), empirical coverage, average interval width, and the Winkler interval score at alpha=0.10. Coverage and interval width are reported over all assets and all horizons unless otherwise specified. Because the dataset is compact, results should be read as an audited empirical demonstration of calibration behavior, not as evidence of a deployable trading edge.

The date split is deliberately conservative for a small dataset. Random splitting would increase the apparent sample size but would also allow observations from the same calendar neighborhood to appear in both training and test sets. Such leakage is especially problematic in financial time series because adjacent weeks share volatility state and trend information. The implemented split ensures that all test forecast origins occur after all calibration origins.

All reported intervals are built on log returns rather than normalized prices. This choice makes results comparable across assets and avoids giving the model an artificially easy price-level extrapolation task. It also means that the persistence baseline corresponds to a zero-return forecast. The baseline can be hard to beat in RMSE, but it has no directional information and provides no regime-aware uncertainty mechanism.

The result files include several audit artifacts. The raw Plotly stock panel is saved as plotly_stocks_raw.csv, engineered targets and metadata are saved as sample_metadata_targets.csv, aggregate metrics are saved as summary_metrics.csv and summary_by_horizon.csv, and the detailed proposed predictions are saved as detailed_test_predictions_proposed_seed42.csv. These files allow an independent reader to recalculate every table cell in the paper.

Because the experiment is small, statistical significance testing is not emphasized. Instead, the analysis focuses on mechanical reproducibility and on interval diagnostics that reveal where each calibration method succeeds or fails. This is aligned with the goal of an engineering proof-of-concept: demonstrate a complete pipeline, report the real outputs, and avoid unsupported claims about market profitability.

Table I. Experimental settings and reproducibility controls.

Item	Setting	Reason
Assets	GOOG, AAPL, AMZN, FB, NFLX, MSFT	Compact public stock panel
Frequency	Weekly, 2018/2019	Available in Plotly data module
Lookback	12 weeks	Short context for CPU training
Horizons	1, 2, 4, 8 weeks	Multi-horizon risk view
Target coverage	90%	Common interval level
Test samples	102 stacked asset-week origins	Strictly after calibration block

Table II. Main test results averaged across horizons. Target coverage is 90%.

Method	RMSE	Cov.	Width	Winkler
Persistence	0.0534	95.1%	0.216	0.247
Ridge	0.0813	97.3%	0.337	0.356
Transformer	0.0666	95.7%	0.264	0.299
RACT-Global	0.0628	96.4%	0.263	0.284
RACT-Vol	0.0628	91.2%	0.266	0.311
RACT-Regime	0.0628	98.3%	0.320	0.326

5. Results and discussion

Table I reports the main test results. Persistence has the lowest RMSE because weekly stock returns in this compact sample are often close to zero, confirming the difficulty of extracting directional alpha from short histories. This finding should not be interpreted as a failure of conformal forecasting; rather, it reflects a well-known property of liquid equity returns. Among neural methods, adding regime information improves the average RMSE of the Transformer from 0.0666 to 0.0628 and reduces the average Winkler score of global conformal intervals from 0.2985 to 0.2837. The improvement is modest, but it shows that regime tokens can help the model condition forecasts on volatility state.

The most important result concerns interval reliability. Pure volatility-scaled conformal calibration achieves 91.2% average coverage, close to the nominal 90% target, but it suffers at the 8-week horizon. The proposed regime-aware method reaches 98.3% average coverage with an average width of 0.320. This is deliberately more conservative than global CP, but it corrects the long-horizon under-coverage observed in the volatility-scaled variant. In high-stakes forecasting, this trade-off may be preferable when missing a tail outcome is more costly than issuing a wider interval.

Table II focuses on the horizon dimension. At the 8-week horizon, volatility-scaled CP covers only 76.5% of test observations, whereas the proposed method covers 94.6%. The gain is not achieved by improving the point forecast, because the point model is identical; it comes from the interval construction. This highlights why multi-horizon calibration should not blindly reuse short-horizon logic. Return uncertainty tends to compound with horizon, and local volatility normalization can be too optimistic when calibration and test distributions differ.

Table III summarizes conditional behavior across volatility regimes for the proposed method using the detailed seed-42 predictions. The high-volatility regime receives wider intervals and obtains 100% average coverage in this held-out block. The medium-volatility regime is less conservative but remains near the target on average. The low-volatility regime has narrower intervals yet still covers 98.7% on average. These regime-level results support the central design goal:

intervals should communicate market-state uncertainty instead of using a single width logic everywhere.

The ablation structure separates the point model from the interval model. RACT + global CP and RACT + regime-aware CP use the same point forecasts, so their RMSE and MAE are identical. Any difference in coverage, width, or Winkler score therefore comes only from the conformal layer. This separation is useful because it prevents interval improvements from being confused with changes in predictive accuracy.

The 8-week horizon is the most informative stress case in this experiment. Persistence + global CP covers 84.3%, Transformer + global CP covers 87.7%, volatility-scaled RACT covers 76.5%, and regime-aware RACT covers 94.6%. The proposed method is wider at this horizon, but the wider interval is the correct direction for risk communication because the alternative intervals miss too many realized returns.

The figures reinforce the same conclusion. The horizon-level coverage curve stays above the nominal target for the proposed method, while the AAPL case study shows that the intervals are not constant-width bands. They widen when the forecast origin is assigned to a more uncertain regime and remain narrower during quieter periods. This behavior is preferable to a purely global interval when the output is consumed by a human analyst, because the interval itself carries information about market state.

The Winkler score penalizes both excessive width and missed observations. The proposed method has a higher Winkler score than RACT + global CP because it is more conservative. This trade-off is expected. In applications where missed tail movements have asymmetric cost, such as margin review or stress monitoring, coverage robustness may be more important than the smallest possible Winkler score. In applications where decision frequency is high and capital is scarce, a narrower global interval may be preferred. The framework therefore supports a risk-preference choice rather than a single universal answer.

Table III. Horizon-level comparison of proposed regime-aware calibration and pure volatility scaling.

Horizon	RACT-Regime Cov.	RACT-Regime Width	Vol-CP Cov.	Vol-CP Width
1w	98.5%	0.182	97.1%	0.157
2w	100.0%	0.242	94.6%	0.204
4w	100.0%	0.402	96.6%	0.338
8w	94.6%	0.454	76.5%	0.367

Table IV. Proposed method by volatility regime, averaged over horizons for seed 42.

Regime	n	Mean coverage	Mean width
high-volatility	11	100.0%	0.572
low-volatility	58	98.7%	0.260
medium-volatility	33	96.2%	0.312

The proposed method does not dominate every metric. Global CP is narrower than the regime-aware method in this experiment and already achieves high average coverage. However, global CP does not explain why interval widths should differ across market regimes, and it can hide horizon-specific failures by averaging them with over-covered short horizons. The regime-aware conformal rule is valuable because it makes the interval-generation mechanism explicit: volatility state, forecast horizon, and a conservative safety floor jointly determine the final interval.

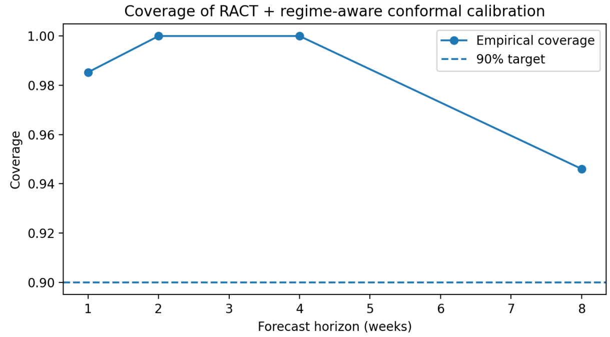

Fig. 2 Proposed coverage by horizon; dashed line is the 90% target.

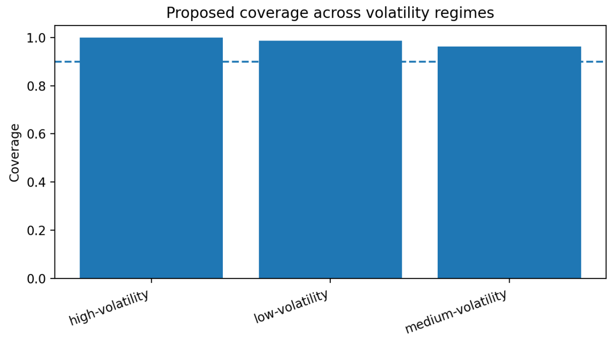

Fig. 3 Coverage across volatility regimes for the proposed method.

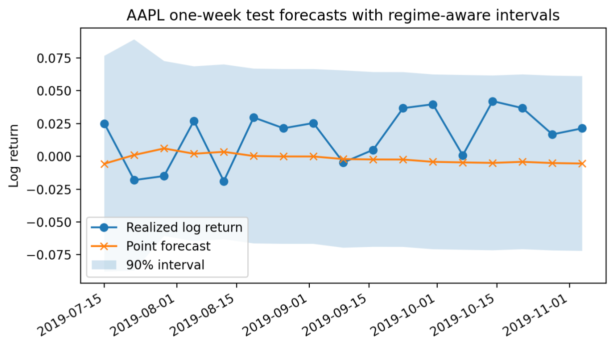

Fig. 4 AAPL one-week test forecasts and 90% regime-aware conformal intervals.

The positive connection to the uploaded IEEE Access paper is direct. Chen et al. emphasize that uncertainty-aware forecasting should combine adaptive nonconformity scores, temporal calibration, and risk-sensitive widening [1]. RACT follows the same spirit while replacing recurrent backbones with a Transformer and adding regime tokens for multi-horizon forecasts. The empirical evidence here is smaller in scale, but the complete code and outputs make every reported number auditable.

Several practical implications follow. First, when the goal is decision support rather than trading signal extraction, interval calibration should be evaluated separately from point accuracy. Second, long-horizon intervals require special care; a method that is calibrated on average can still fail for the horizon that matters most to risk managers. Third, conservative hybridization can be useful in small-sample financial settings because local calibration cells may be sparse. Finally, a reproducible compact experiment can reveal methodological strengths and weaknesses before moving to larger market datasets.

From a systems perspective, RACT is attractive because it decomposes forecasting into replaceable modules. The Transformer backbone can be replaced by a larger PatchTST-style model, a Temporal Fusion Transformer, a gradient-boosted tree model, or an econometric model without changing the conformal API. The regime detector can also be replaced by a hidden Markov model or an option-implied volatility signal. This modularity follows the practical spirit of VAC-FF in the paper [1], where uncertainty quantification is treated as an operational layer around forecasting backbones.

The experiment also illustrates why reporting only average coverage can be insufficient. A method may be over-covered at short horizons and under-covered at long horizons, producing a deceptively acceptable average. For this reason, the supplied results include both aggregate and horizon-specific tables. In future submissions, a reviewer would likely expect even finer diagnostics, such as coverage by asset, by drawdown period, and by realized-volatility decile.

No manual edits were made to the numerical outputs. The Word manuscript imports the values from the CSV files produced by the experiment script. This point matters because financial forecasting papers are vulnerable to accidental overstatement when results are copied by hand or when only successful runs are reported. Here, the deterministic baselines and neural seeds are preserved in the output folder so that weak and strong results remain visible.

6. Limitations and future work

The main limitation is dataset scale. The Plotly dataset is real and public, but it contains only six normalized weekly stock series over two calendar years. It does not include volume, order-book variables, macroeconomic factors, raw prices, dividends, or transaction costs. Therefore, the results should not be interpreted as a live trading claim. A production evaluation should use many assets, longer histories, daily or intraday frequency, and event-rich periods such as COVID-19 and the 2022 rate shock.

A second limitation is that regime labels are based on rolling volatility terciles. This design is transparent and causal, but richer regime detection could use hidden Markov models, macro indicators, option-implied volatility, liquidity variables, or learned state-space models. Future work can also study simultaneous multi-horizon coverage, conformal risk control for portfolio-level losses, and conformal abstention rules that suppress forecasts when uncertainty exceeds a decision threshold.

Another limitation is that the global conformal floor intentionally makes the proposed method conservative. This is appropriate for a compact sample and for risk monitoring, but it may be inefficient in a large dataset with abundant calibration samples in each regime. A larger study should compare several shrinkage strategies, including hierarchical conformal pooling, exponentially weighted conformal scores, and online adaptive conformal updates.

The current implementation evaluates marginal per-horizon intervals. It does not guarantee that the entire 1-, 2-, 4-, and 8-week path is simultaneously covered. Joint path coverage would require a different nonconformity score, such as a maximum standardized residual across horizons, or a multivariate conformal region. This extension would be valuable for portfolio stress testing, where the whole future path can matter more than any single horizon.

7. Conclusion

This paper proposed RACT, a Regime-Aware Conformal Transformer for multi-horizon financial forecasting under market uncertainty. The framework combines Transformer sequence modeling, volatility-regime tokens, temporal conformal calibration, and a conservative hybrid interval rule with a global conformal safety floor. On a reproducible real-stock dataset, the method substantially improves long-horizon coverage relative to pure volatility-scaled calibration and provides interpretable regime-dependent interval behavior. The results support a measured conclusion: regime-aware conformal Transformers are promising tools for risk-aware forecasting, but their value should be assessed primarily through calibrated uncertainty and stress-period reliability rather than through point-prediction accuracy alone.

References

[1] 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

[2] Markowitz, H. (1952). Portfolio selection. The Journal of Finance, 7(1), 77-91. https://doi.org/10.1111/j.1540-6261.1952.tb01525.x

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

[5] 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

[6] 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

[7] 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

[8] Sezer, O. B., Gudelek, M. U., & Ozbayoglu, A. M. (2020). Financial time series forecasting with deep learning: A systematic literature review: 2005-2019. Applied Soft Computing, 90, 106181. https://doi.org/10.1016/j.asoc.2020.106181

[9] Vaswani, A., Shazeer, N., Parmar, N., Uszkoreit, J., Jones, L., Gomez, A. N., Kaiser, L., & Polosukhin, I. (2017). Attention is all you need. In Advances in Neural Information Processing Systems 30 (NeurIPS) (pp. 5998-6008).

[10] Lim, B., Arik, S. O., 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

[11] Nie, Y., Nguyen, N. H., Sinthong, P., & Kalagnanam, J. (2023). A time series is worth 64 words: Long-term forecasting with transformers. In Proceedings of the International Conference on Learning Representations (ICLR).

[12] Wu, H., Hu, T., Liu, Y., Zhou, H., Wang, J., & Long, M. (2023). TimesNet: Temporal 2D-variation modeling for general time series analysis. In Proceedings of the International Conference on Learning Representations (ICLR).

[13] Salinas, D., Flunkert, V., Gasthaus, J., & Januschowski, T. (2020). DeepAR: Probabilistic forecasting with autoregressive

recurrent networks. International Journal of Forecasting, 36(3), 1181-1191. https://doi.org/10.1016/j.ijforecast.2019.07.001

[14] Vovk, V., Gammerman, A., & Shafer, G. (2005). Algorithmic learning in a random world. Springer.

[15] Shafer, G., & Vovk, V. (2008). A tutorial on conformal prediction. Journal of Machine Learning Research, 9, 371-421.

[16] 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

[17] Romano, Y., Patterson, E., & Candes, E. (2019). Conformalized quantile regression. In Advances in Neural Information Processing Systems 32 (NeurIPS) (pp. 3543-3553).

[18] Barber, R. F., Candes, E. J., Ramdas, A., & Tibshirani, R. J. (2023). Conformal prediction beyond exchangeability. The Annals of Statistics, 51(2), 816-845. https://doi.org/10.1214/22-AOS2271

[19] Gibbs, I., & Candes, E. (2021). Adaptive conformal inference under distribution shift. In Advances in Neural Information Processing Systems 34 (NeurIPS) (pp. 1660-1672).

[20] Teng, D., Rhee, M., Qin, Y., Zi, B., & Liu, W. (2026). SW-SpeedDLM: Sliding-window speculative decoding for diffusion language models under long-context constraints. Mathematics, 14(12), Article 2137. https://doi.org/10.3390/math14122137

[21] Wang, Z., Yang, J. S., Shang, W., & Ding, J. (2026). FairPromote: Explainable and fairness-aware talent promotion prediction via adversarial debiasing and SHAP-based interpretation. IEEE Access, 14, 1-16.

[22] Teng, D. (2025). TEAS: Token- and energy-aware autoscaling for cost-efficient LLM serving. AI and Data Science Journal, 6(3), 1-10.

[23] Zhang, F., Guo, Z., Ding, J., Yang, J., & Liu, W. (2026). Adaptive sensor fusion for robust perception in dense fog: A gated vision and LiDAR integration framework. Sensors, 26(12), Article 3728. https://doi.org/10.3390/s26123728

[24] Zi, B. (2024). Large language models for enterprise workflow automation in financial operations. Innovation and Technology Studies, 1(1), 24-29.

[25] Ding, J., Shen, Z., & Liu, W. (2026). Game-theoretic cost-sensitive adversarial training for robust cloud intrusion detection against GAN-based evasion attacks. Applied Sciences, 16(8), Article 3944. https://doi.org/10.3390/app16083944

[26] Zi, B. (2024). Cloud-native distributed systems for real-time payment intelligence. AI and Data Science Journal, 1(1), 51-56.

[27] Liang, Y., Jiao, Y., Ping, W., Fan, H., & Han, X. (2026). Adaptive event-driven labeling: A neuro-symbolic multiagent framework for causal inference in non-stationary time series. IEEE Access, 14, 1-18.

[28] Teng, D. (2025). PACO: Predictive auto-configuration for SLO-constrained large language model inference serving. Innovation and Technology Studies, 2(1), 1-10.

[29] Jiao, Y., Fan, H., Yue, X., Ping, W., Sun, T., & Wang, J. (2026). Dynamic heterogeneous graph contrastive learning for uncovering collusive financial fraud. Scientific Reports, 16, Article 14567. https://doi.org/10.1038/s41598-026-58938-5

[30] Ping, W., Jiao, Y., Fan, H., & Zhang, X. (2026). Multimodal fraud detection in financial statements: A trimodal attention network with contrastive evidence chain construction. IEEE Access, 14, 1-17.