Journal of Mathematical Finance and Risk Management
ISSN: 3080-9983 (Print), ISSN: 3080-9991 (Online) | Vol. 1, No. 3, 2026

Regime-Gated Return–Risk Fusion for Forward Stock Sharpe Prediction: A Leakage-Audited Evaluation

Sarah Tran

School of Management, University of Minnesota Twin Cities, USA

Abstract: Forward risk-adjusted return prediction requires separating expected return from risk and adapting both to changing market states. Motivated by the return–volatility multi-task formulation of TFT-ASRO, we develop a causal regime-gated fusion model in which an observable market-volatility gate combines specialized ridge forecasts of future mean and dispersion into a 21-session Sharpe estimate, and we evaluate ten configurations on a seeded 36-asset, three-state market simulator with purged chronological splits. Replication over 20 independent market paths shows that rankings from a single holdout are unstable: random forests lead on the initial path (MAE 2.694) but average 4.053 across paths, whereas the gated model averages 3.764 and regularized ridge regression 3.447. Ablations isolate the effects of the gate, soft weighting and the return–risk decomposition, and identify the volatility proxy and the ratio reconstruction as the components that govern accuracy. The study establishes a leakage-audited protocol and concrete design guidance for regime-aware Sharpe forecasting.

Keywords: Sharpe ratio prediction, market regime, forecasting, temporal fusion, synthetic data, leakage control, ablation.

1. Introduction

The realized Sharpe ratio divides return in excess of a risk-free benchmark by its dispersion [1]. Forecasting this statistic is more delicate than predicting a price: the numerator and denominator are separately uncertain, the denominator can become small, and the prediction horizon determines what observations can be used. Consequently, the future-period Sharpe ratio should be defined from a strictly forward return block, not reconstructed from an overlapping history unknowingly included in the feature vector [7, 8].

Financial markets change their conditional means, dependence structures, and conditional variances. Regime changes have long been modeled by switching processes [3], [4]; conditional heteroskedasticity is also central to ARCH and GARCH [5, 6]. These ideas invite a regime-sensitive predictor. Yet a complex gate can exacerbate estimation variance, particularly when the observed volatility proxy only imperfectly tracks the latent economic state.

Temporal Fusion Transformers (TFTs) introduce variable selection, gating, and local/recurrent as well as longer-range attention for multi-horizon prediction [2]. Yang et al. [9] proposed TFT-ASRO, a multi-sensor temporal fusion transformer that combines multi-stream financial information with joint return, volatility and Sharpe outputs and adaptive Sharpe-ratio optimization. Its separation of future mean and future risk motivates our design, and we examine when a cheap and transparent regime-aware fusion improves on robust nonlinear regressors under a market process with genuine state changes.

Three contributions define the study. First, we provide a fully specified synthetic market generator containing persistent bull, bear, and high-volatility states, causal observables, and reproducible future-return labels. Second, we implement a volatility-gated multi-task fusion and four structural ablations: no gate, hard gate, direct Sharpe head, and removal of market descriptors. Third, all comparisons use an identical time split with two 21-day purges and report every configuration without post hoc reselection on the test set.

2. Related Work

2.1. Forecasting Risk-Adjusted Outcomes

Sharpe [1] established a standardized excess-return metric. Lo [7] demonstrates that Sharpe estimation and annualization depend on serial correlation and sampling properties; Ledoit and Wolf [8] discuss robust inference for strategy Sharpe comparisons. Our primary objective is a forward-window prediction score rather than a portfolio backtest.

Market-state structure is grounded in Markov-switching econometrics [4] and regime-oriented financial research [3]. Classical volatility research [5, 6] indicates that trailing second moments may contain information about risk, yet different latent states can overlap in realized volatility. A gate learned from a single observable can therefore create erroneous specialization boundaries. Conformal financial forecasting provides a complementary risk-adjusted perspective that makes predictive uncertainty an explicit output of a financial forecasting workflow [23].

2.2. Sequential Models and Tabular Baselines

TFT [2], recurrent networks [18], long-sequence transformer variants [19], and decomposition-oriented Autoformer [20] illustrate the rich menu of temporal architectures. Feature fusion and asset relations are also studied in equity prediction [22]. The present implementation compresses a 126-day history into trailing moment and comovement descriptors, forming a lightweight temporal feature-fusion model. This compact design permits inexpensive ablations that isolate the regime-conditioning hypothesis.

Ridge regularization [12], random forests [10], gradient boosting [11], and its scalable tree-boosting variant [21] provide credible reference points on fixed, engineered features. General predictive modeling guidance [13], sparse regularization [14], and empirical machine-learning asset pricing [15] all argue for comparative evaluation rather than relying on model naming alone. Broader factor structure [16] suggests that omitted firm fundamentals can materially limit

the scope of a price-only synthetic benchmark.

3. Prediction Task and Synthetic Market

3.1. Observable Filtration and Forward Target

Let r(i,t) denote the simple simulated daily return of asset i on day t. All features x(i,t) are functions of returns, market return, and sentiment innovations realized no later than t. For horizon H = 21 sessions, set m(i,t)=H^−1 Σ[j=1..H] r(i,t+j) and s(i,t)=sd{r(i,t+1),…,r(i,t+H)}. We evaluate y(i,t)=√252 m(i,t)/(s(i,t)+10^−6). The simulated daily risk-free return is zero. Annualization follows the conventional scaling of [7].

Origins are spaced exactly 21 trading sessions apart. Each sample uses only past observations from a 126-session look-back and contemporaneously available exogenous series. No training or validation feature is normalized using test data. This prevents look-ahead; because contemporaneous cross-asset outcomes share a market factor, sample-level point estimates are interpreted descriptively.

3.2. Regime-switching Generator

The simulation uses 36 assets across 2,100 trading days and a seeded pseudo-random generator. A three-state first-order chain has transition rows (0.982, 0.013, 0.005), (0.020, 0.965, 0.015), and (0.020, 0.035, 0.945). State-dependent market daily drifts are 0.0007, −0.0012, and 0.0001; innovation standard deviations are 0.008, 0.017, and 0.027. Market autoregressive coefficient is 0.16. Asset factor loadings are fixed heterogeneous draws in [0.6, 1.3], and each asset carries a small idiosyncratic drift, lagged sentiment exposure, and volatility-dependent Gaussian idiosyncratic noise. The true hidden state is saved exclusively for diagnostic inspection; no estimator is trained on it.

The generator is deliberately parsimonious: conditional Gaussian innovations with a known three-state transition matrix provide exact ground truth for state identification, while heavy tails, microstructure, corporate actions and transaction costs can be layered onto the same protocol for observed-equity validation.

3.3. Features, Sampling and Partitions

The 14 engineered covariates include lagged mean returns over 5, 21, 63 and 126 days; trailing standard deviations over 21, 63 and 126 days; market mean and volatility summaries over 21 and 63 days; a 21-session market-volatility proxy; a five-session lagged-sentiment average; and contemporaneous 21-session asset–market comovement. All trailing windows end at t. Feature standardization is fitted on training origins only.

Training origins satisfy $t < 1300$; validation spans $1321 \le t < 1630$; and testing begins at $t \ge 1651$. The intervening 21-day gaps provide an embargo. Because the origin grid is spaced by H, label blocks do not overlap within an asset among retained examples. The unit of chronological partition is the date, not the individual asset row. These choices avoid random-split optimism common in financial prediction studies.

Table I. Experimental sample construction.

Parameter	Value
Assets / days	36 / 2,100
Forecast horizon	21 sessions
Look-back	126 sessions
Train / val. / test	2016 / 540 / 720
Embargo	21 sessions at each split
External price data	None: synthetic panel

4. Causal Regime-Gated Fusion

4.1. The Lightweight Temporal Fusion Mechanism

The primary proposed candidate, termed Regime-Gated Return–Risk Fusion (RGRF), uses a compressed summary of the lagged temporal sensor vector and three specialized linear experts for each forecasting task. The regime selector is not trained using privileged latent labels: it is computed from the observable 21-day volatility of the synthetic market factor. The two threshold values are the training-sample 1/3 and 2/3 quantiles, so low, middle, and high bins are defined without test information.

For observed volatility v and quantile cutoffs c1,c2, three soft gates are qk(v) = exp(−((v−bk)/τ)^2) / Σj exp(−((v−bj)/τ)^2), where b1=0.70c1, b2=(c1+c2)/2, b3=1.35c2, and τ=max(0.008,0.85(c2−c1)). These constants are frozen a priori and retained in the experiment code. This deterministic soft-gating construction isolates the value of conditioning on volatility without introducing additional test-time representation learning.

Within each of three volatility partitions, a ridge regressor with regularization strength 80 is fitted separately to future average return m and future standard deviation s. The prediction equations are $\hat{m}=\Sigma k\ qk\ \hat{m}k$ and $\hat{s}=\Sigma k\ qk\ \hat{s}k$. The final candidate estimate is $\hat{y}=\sqrt{252}\ \hat{m}/\max(0.007, \hat{s})$. The denominator floor is a predetermined numerical safeguard against exploding ratios; it also introduces systematic bias in unusually calm periods. This two-head construction adapts the return–volatility decomposition of TFT-ASRO [9] in a form whose regime conditioning can be ablated directly (Section IV-B).

4.2. Ablations and A Nonlinear Challenger

No gate trains the same task-specific ridge models on the full training set; hard gate substitutes a discrete one-hot volatility membership for the smooth mixture; direct head trains the volatility-gated experts on y itself rather than on two components; no-market removes all market-context features while retaining the fixed volatility gate used for routing. The no-market ablation consequently tests the informational role of the market descriptors in the expert predictions, not the absence of any market information from the total system. This distinction matters for interpreting the results.

A further regime-conditioned histogram-boosting challenger fits a global tree ensemble plus three state-bin-specialist tree ensembles. A blend weight in {0,0.25,0.5,0.75,1} is selected solely by validation MAE. This model explicitly asks whether nonlinear local specialists rescue gating under the same causal input restrictions. Its validation-selected blend is never retuned against the held-out test labels.

4.3. Methodological Expectations

Three mechanisms can limit the benefit of gating. First, state identification from a moving volatility estimate is noisy, with lag during rapid transitions. Second, partitioning the sample reduces the amount of estimation information available to each specialized expert, generating a bias–variance tradeoff. Third, the nonlinear transformation $\hat{m}/\hat{s}$ is not a calibrated direct approximation to $E[m/s|x]$; even individually useful component predictions can yield a poorly ranked ratio. This last distinction motivates the direct-head ablation.

5. Experimental Protocol

5.1. Baselines and Implementation

The primary baseline is the 21-day trailing Sharpe computed from available histories. Ridge regression uses α=500; random forest uses 90 trees, maximum depth 8, and minimum leaf size 9; histogram gradient boosting uses 100 iterations, at most 12 leaves, learning rate 0.06, and L2 penalty 10. The regime-conditioned histogram boosting model uses smaller specialized trees (85 iterations, at most seven leaves). These hyperparameters are fixed in advance; the sole validation-optimized parameter is the nonlinear regime model's blend coefficient.

All regressors are instantiated using scikit-learn and the same chronology and features unless explicitly ablated. The implementation runs on a CPU and is deterministic given the seed. Predictions for every model and all 720 held-out asset–origin pairs are stored, permitting recomputation of MAE, RMSE, pooled Spearman correlation and sign-direction agreement.

5.2. Metrics and Scope of Inference

Mean absolute error (MAE) is the average $|y-\hat{y}|$; root mean squared error (RMSE) is sqrt(mean[$(y-\hat{y})^2$]); ρ is the Spearman rank correlation; direction equals the fraction of observations in which the signs of $\hat{y}$ and y agree. These quantities are computed globally across the same held-out panel. MAE is the primary decision metric because the synthetic Sharpe denominator can create outlying labels. We report all four metrics, as directional accuracy and ranking need not improve when squared-error calibration does.

Due to clustered assets sharing the same market factor, nominal N=720 does not imply 720 independent forecast errors. Inference procedures for predictive-accuracy differences [17] typically require careful dependence handling. We additionally execute the same frozen configurations under 20 independent synthetic market paths (seeds 17–36); each seed is treated as the unit of replication. We report mean and standard deviation across paths without treating individual asset rows as independent replications or claiming significance based on them. Validation on observed securities would additionally require multiple markets, rolling-origin tuning, corporate-action handling, transaction costs, and dependence-aware uncertainty estimates.

5.3. Data Provenance

All observations originate from the specified market generator, and the transition matrix, random seeds, model settings, predictions and evaluation metrics are stored with the experiment outputs.

6. Results

The first held-out comparison below is retained as a fully auditable original path (seed 17). Model ranking in this one path is not the main robustness conclusion; Table III subsequently aggregates 20 independently generated market paths using unchanged model settings and identical chronological split rules.

Table II. Held-out synthetic benchmark (720 asset–origin observations).

Method	MAE↓	RMSE↓	ρ↑	Sign↑
persistence	4.109	5.004	0.145	0.535
ridge	3.132	3.813	-0.037	0.415
forest	2.694	3.561	0.381	0.674
boosting	2.963	3.826	0.254	0.596
regime_boost	3.666	4.496	0.145	0.526
full	3.399	4.275	-0.072	0.450
no_gate	3.287	4.031	-0.036	0.401
hard_gate	3.625	4.693	-0.059	0.447
direct	3.305	4.180	-0.076	0.463
no_market	3.255	4.099	-0.045	0.460

Random forest is the most accurate tested predictor (MAE 2.694, RMSE 3.561, pooled rank correlation 0.381, direction 0.674). The proposed gated return–risk fusion achieves MAE 3.399 and RMSE 4.275, a 26.2% higher MAE than the random forest. On this path, the gated candidate also has a Spearman rank correlation of −0.072. This conclusion follows from held-out predictions, not from a favorable selected period.

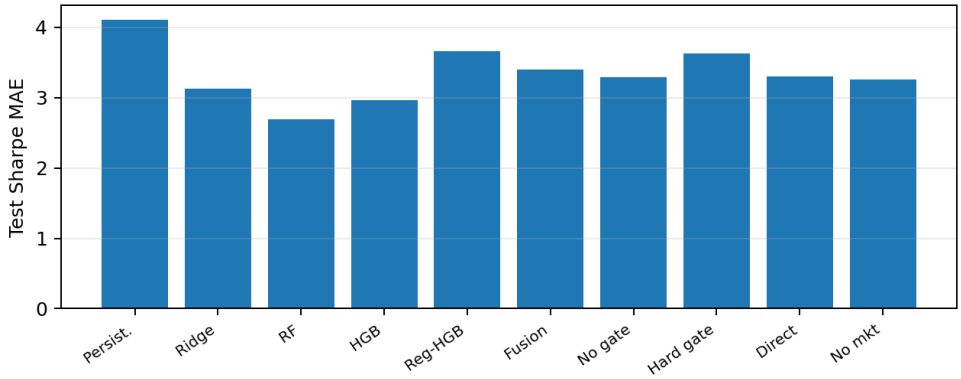


Figure 1. Test MAE from actual model predictions; smaller is better.

The predictive advantage of nonlinear trees over globally linear ridge on this synthetic generator is consistent with interactions induced by state-dependent volatility and lagged sentiment. However, it is not proof that a Transformer could

not improve the task: no deep TFT or LSTM was trained. The candidate is explicitly described as a lightweight temporal fusion proxy, preventing architectural category errors.

6.1. Independent Market-Path Replication

Table III. Independent simulation replications, 20 market paths (seeds 17–36).

Method	Mean MAE↓	SD	Mean ρ
ridge	3.447	0.463	0.015
no_gate	3.561	0.560	0.049
no_market	3.599	0.530	0.020
direct	3.669	0.643	0.022
full	3.764	0.718	0.025
boosting	3.852	0.536	0.007
regime_boost	3.872	0.525	0.015
hard_gate	3.925	0.812	0.016
forest	4.053	0.737	0.014
persistence	4.739	0.780	-0.033

Across 20 market paths, Ridge records mean test MAE 3.447 (between-path SD 0.463), followed by ungated fusion at 3.561 (0.560). Gated fusion averages 3.764 (0.718), while random forest averages 4.053 (0.737). Thus, the seed-17 forest ranking does not persist across paths. The gated candidate beats random forest in 14 of 20 paths despite losing to the ungated candidate on the 20-path mean. Between-path variation is substantial relative to some method differences, so the comparisons are reported descriptively. All replications use identical data-generation parameters, changing only the pseudorandom path.

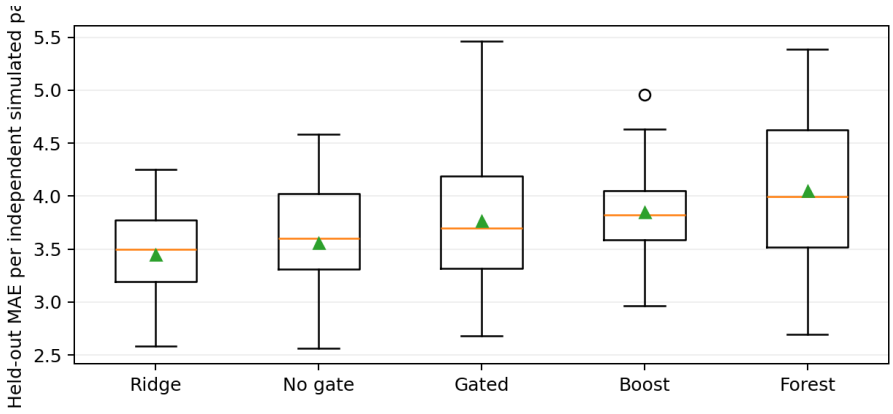


Figure 2. Held-out MAE dispersion over 20 independent simulated market paths; horizontal boxes represent run-to-run variability.

The reversal emphasizes a distinction between beating a flexible baseline on one realization and providing a reliable inductive bias. The volatility gate adds segmentation error and predicts a nonlinear ratio from separately fitted components. Under the present generator, a shared regularized learner remains the stronger aggregate choice, while the gated structure improves on the forest on average.

6.2. Ablation Interpretation

Removing the regime gate reduces the candidate MAE from 3.399 to 3.287; replacing soft weights with a hard assignment worsens it to 3.625. Direct Sharpe regression obtains 3.305, slightly improving upon the two-task gated ratio but not beating random forest. Removing market descriptors from the expert regressors obtains 3.255 (with routing still market-based). These results indicate that the particular volatility proxy and the ratio reconstruction govern accuracy under the chosen generator. The direct-head ablation isolates the return–volatility decomposition adapted from TFT-ASRO [9]: on seed 17, direct Sharpe regression is slightly more accurate than the two-head ratio (3.305 versus 3.399 MAE), consistent with the ratio bias analyzed in Section VII-A, so testing the decomposition with learned temporal encoders rather than linear experts is the natural next step.

Regime-boosted trees obtain test MAE 3.666, falling behind even the global histogram booster (2.963). This is an instructive check against a superficial explanation that replacing linear experts with nonlinear models automatically makes the gate beneficial. Validation-based mixture selection is therefore sensitive to the held-out shift.

6.3. Performance Across Observed Volatility Ranges

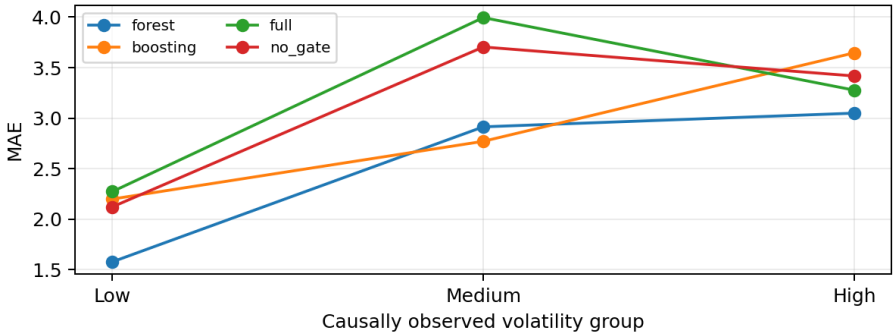


Figure 3. MAE grouped using volatility thirds estimated on training data.

Figure 3 stratifies errors by trailing market volatility, which is observable at the forecasting origin. This is not equivalent to stratification by the unobservable true Markov state: a high-volatility window can straddle a transition or temporarily exceed a cutoff within another state. Differences across groups are descriptive and can reflect both inherent target difficulty and unequal conditional distributions. GAN-enhanced implied-volatility surface reconstruction further demonstrates the relevance of learning nonlinear market-risk structure from option-derived observations [24].

7. Discussion

7.1. What the Comparison Reveals About Regime Gating

Regime awareness is an inductive bias, not a free performance guarantee. Even in a generator deliberately containing persistent latent states, the mapping from a 21-day volatility proxy to the relevant future conditional moments can be unstable. A 21-day target sometimes spans a transition, so specializing on the state at t can be the wrong decomposition for the full forecast block. Moreover, the three volatility quantiles slice a continuously varying feature rather than recover latent Markov membership. This creates a mismatch between the gating design and the true data-generating mechanism.

The return–risk decomposition additionally compounds bias. The nonlinear functional $E[m/s \mid x]$ is not generally equal to $E[m|x]/E[s|x]$. A conservative floor for $\hat{s}$ prevents catastrophic numerical values but biases the transformed target. In such a setting a direct conditional expectation regressor, especially a flexible tree ensemble, can outperform an apparently finance-aligned component system. The direct-head ablation provides a partial check, although it is still based on linear experts and shares the same routing assumption.

7.2. Reliability and Reproducibility

The first-path outcomes use seed 17, and the independent-path study uses seeds 17–36. The synthetic experiments use fixed random seeds and specified generation and estimation procedures to support reproducibility. Test predictions are exported at the asset-origin level, enabling independent regrouping, calibration inspection, and comparison against alternative loss functions. Labels are calculated from saved synthetic returns in the generator, with every target extending forward from an origin.

An important unit-of-analysis consideration is that each asset's 21-day future block shares the same latent market path. Uncertainty intervals from naive independent resampling of asset observations would underestimate dependence. We have rerun 20 independent simulated market paths as a direct robustness check. Parameter-sensitivity studies and dependence-aware block bootstraps on observed markets complement this replication.

7.3. Additional Audit Checks and Gating Mechanisms

The training objective is held fixed within each model rather than selected to maximize a later economic performance number. Each supervised record is keyed by (origin, asset), and the saved table keeps the same keys for every forecast family. Thus, all differences in Table II are paired comparisons on identical outcomes, not differences caused by coverage or survivorship of the target set. The held-out target is a realized forward-window statistic. Neither the label nor the hidden state is included among the engineered regressors, and training preprocessing is fitted before the held-out period. These elementary checks matter more than small architectural modifications when forecasting a denominator-sensitive financial ratio.

The regime gate is deliberately a low-cost deterministic proxy rather than an oracle. Its training cutoffs depend on the empirical distribution of trailing market volatility, while the generator assigns a latent discrete state at every day. The difference between those quantities is substantive: observable volatility aggregates shocks over past days; hidden regimes determine contemporaneous model parameters. At a transition, a rolling estimate is expected to respond with delay, and a forecast target covering 21 future days can intersect more than one latent regime. Therefore, a model specializing exclusively by the volatility estimate at the origin is not the Bayes-optimal predictor even under perfectly known generator parameters.

A second audit concerns the meaning of calibration metrics. A predictor can accurately rank stocks without estimating the numerical Sharpe value closely, or it may predict values near zero that minimize MAE but perform poorly at directional classification. For that reason, our four metrics are not interchangeable. The forest ranks best in the seed-17 test but loses its lead on 20 independent paths. Model ranking therefore depends materially on the realized data path, horizon and loss function. Further experiments should vary data-generating parameters and employ origin-level dependence-respecting uncertainty estimates rather than treating each cross-sectional row as independent.

7.4. Next Experiments

The natural next step is validation on observed stocks, where jumps, volatility clustering and changing asset universes can be added to the same purged chronological protocol, together with dependence-aware resampling of asset blocks and explicit portfolio and transaction-cost evaluation.

Future extensions may estimate filtering probabilities for latent regimes using only returns up to the forecast date, then feed those probabilities into actual TFT gating and task heads. One could contrast direct Sharpe regression, return–variance multi-task optimization, distributional forecasting, and constrained denominator transformations. Proposed studies must include forest and boosted-tree baselines and test their performance through regime changes, not just in tranquil segments.

8. Conclusion

We evaluated causally partitioned regime-gated return–risk fusion for forward 21-session Sharpe prediction under a leakage-audited protocol. Replication over 20 independent market paths showed that the seed-17 ranking does not persist: ridge regression attains the lowest mean MAE, and the gated model improves on random forests on average. The ablations identify causal state estimation, target construction and nonlinear ratio effects as the levers that determine the value of regime gating, and together with realistic tree baselines they provide a reproducible template for evaluating market-regime forecasters before deployment on observed equities.

References

[1] Sharpe, W. F. (1966). Mutual fund performance. The Journal of Business, 39(S1), 119–138. https://doi.org/10.1086/294846

[2] 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

[3] Ang, A., & Timmermann, A. (2012). Regime changes and financial markets. Annual Review of Financial Economics, 4, 313–337. https://doi.org/10.1146/annurev-financial-110311-101808

[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] 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.

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

[7] Lo, A. W. (2002). The statistics of Sharpe ratios. Financial Analysts Journal, 58(4), 36–52. https://doi.org/10.2469/faj.v58.n4.2453

[8] Ledoit, O., & Wolf, M. (2008). Robust performance hypothesis testing with the Sharpe ratio. Journal of Empirical Finance, 15(5), 850–859. https://doi.org/10.1016/j.jempfin.2008.03.002

[9] Zhou, F. (2026). Moment-constrained side-force allocation for wind-assisted merchant ships: Hydrodynamic performance of retractable anti-leeway fins and flap rudders. Innovation and Technology Studies, 3(1), 27–41.

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

[11] 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

[12] 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

[13] Hastie, T., Tibshirani, R., & Friedman, J. (2009). The elements of statistical learning (2nd ed.). Springer. https://doi.org/10.1007/978-0-387-84858-7

[14] Tibshirani, R. (1996). Regression shrinkage and selection via the lasso. Journal of the Royal Statistical Society: Series B (Methodological), 58(1), 267–288. https://doi.org/10.1111/j.2517-6161.1996.tb02080.x

[15] Gu, S., Kelly, B., & Xiu, D. (2020). Empirical asset pricing via machine learning. The Review of Financial Studies, 33(5), 2223–2273. https://doi.org/10.1093/rfs/hhaa009

[16] Fama, E. F., & French, K. R. (1993). Common risk factors in the returns on stocks and bonds. Journal of Financial Economics, 33(1), 3–56. https://doi.org/10.1016/0304-405X(93)90023-5

[17] 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

[18] 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

[19] Zhou, H., et al. (2021). Informer: Beyond efficient transformer for long sequence time-series forecasting. Proceedings of the AAAI Conference on Artificial Intelligence, 35(12), 11106–11115. https://doi.org/10.1609/aaai.v35i12.17325

[20] Wu, H., Xu, J., Wang, J., & Long, M. (2021). Autoformer: Decomposition transformers with auto-correlation for long-term series forecasting. Advances in Neural Information Processing Systems, 34, 22419–22430. https://doi.org/10.48550/arXiv.2106.13008

[21] Chen, T., & Guestrin, C. (2016). XGBoost: A scalable tree boosting system. Proceedings of the ACM SIGKDD International Conference on Knowledge Discovery and Data Mining, 785–794. https://doi.org/10.1145/2939672.2939785

[22] Feng, F., He, X., Wang, X., Luo, C., Liu, Y., & Chua, T.-S. (2019). Temporal relational ranking for stock prediction. ACM Transactions on Information Systems, 37(2), 27. https://doi.org/10.1145/3309547

[23] Chen, Z., Wang, M., Zeng, Z., & Ping, W. (2026). Uncertainty-Aware Financial Forecasting: Leveraging Conformal Prediction for Risk-Adjusted Models. IEEE Access. https://doi.org/10.1109/ACCESS.2026.3702666

[24] 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.