Proceedings of the 8th International Conference on Computing and Data Science
DOI: 10.54254/2755-2721/2026.37873

Regime-Adaptive Multi-Scale Random Fourier Features for Stock Price Forecasting

Yan Guan

Jinan University-Birmingham University Joint Institute, Jinan University, Guangzhou, China
2452736491@qq.com

Abstract. Stock price forecasting is fundamentally difficult because of market non-stationarity, high volatility, and the weak predictability of intraday returns. While kernel methods with random features provide a theoretically grounded framework for capturing non-linear relationships, they are typically applied with a single, globally fixed feature representation that ignores the fact that the generative process of prices varies across market regimes. This paper proposes the Regime-Adaptive Multi-Scale Random Fourier Features (RA-MSRFF) model, which combines two ideas: (i) a multi-scale Random Fourier Feature (RFF) mapping that approximates multiple radial-basis kernels at different bandwidths and feeds a ridge regressor, and (ii) a soft regime-gating mechanism in which a volatility-driven fuzzy membership decomposes each sample into overlapping regimes, each equipped with its own local ridge expert, and the final prediction is a convex fusion of the global multi-scale estimator and the regime-aware local estimators. We establish that the regime-mixed kernel remains positive definite, that the fused objective stays convex, and that soft gating controls regime-misassignment risk. On two representative A-share bank stocks (601166, 601398) using a strict chronological train/validation/test protocol, RA-MSRFF strictly dominates five baselines—Open-price anchor, Static MS-RFF, KNN, AdaBoost, and Previous-close—on RMSE, MAE, MAPE, and R^2.

Keywords: Stock price forecasting, Random Fourier Features, kernel methods, regime adaptation, fuzzy membership, ridge regression, soft gating

1. Introduction

Forecasting the closing price from information available at the market open is a deliberately narrow problem. The open price already summarizes overnight information and, for liquid bank shares, usually lies close to that day's close. A useful model must therefore explain a small open-to-close residual rather than simply reproduce the price level. That residual is noisy, and its behavior does not remain constant through time. For example, a lagged return or volume signal observed during a quiet period may have a different effect after volatility rises. A model fitted mostly on tranquil observations can then be pulled off course by a short turbulent interval. This is why a single relationship estimated over the full sample is questionable. It is also why the open price is a demanding benchmark: a more elaborate method is useful only if it improves on that anchor under a

© 2026 The Authors. This is an open access article distributed under the terms of the Creative Commons Attribution License 4.0 (https://creativecommons.org/licenses/by/4.0/).

DOI: 10.54254/2755-2721/2026.37873

chronological, leakage-free evaluation. We examine this comparison for two liquid A-share bank stocks.

Earlier studies point to several possible model families, but they do not give a settled answer for this particular forecasting setup. A generalized linear model has been applied to close-price prediction for banking stocks [1], offering a transparent global relation between inputs and the response. K-nearest-neighbor (KNN) regression removes that parametric restriction: an optimized small-sample KNN procedure is studied in [2], and a separate structured prediction application is reported in [3]. Broader comparisons of supervised learners also retain KNN and support vector machines as accessible nonlinear benchmarks [4]. Stock-oriented deep models take the opposite route and add substantial temporal capacity. Examples include an LSTM tuned by an artificial-rabbits optimizer [5] and a moving-average-assisted deep architecture [6]. A pretrained model for chaotic dynamics further shows how synthetic training can support zero-shot forecasts [7]. These methods are expressive, although their data and tuning requirements are not always a good match for a two-stock tabular study. Random features provide a lighter alternative. Fourier embeddings can encode continuous numerical inputs compactly [8], while random feature mapping has also been combined with AdaBoost-style result fusion [9]. Combination alone, however, does not remove the effect of noisy or overlapping distributions; weak AdaBoost outcomes under such conditions are documented in [10]. The applications in [2-4, 8, 9] are not identical to stock forecasting, so here they are used to motivate algorithmic choices rather than as direct performance comparisons. Read together, this literature leaves a practical gap: the model should remain small enough for ridge-type estimation, capture nonlinear structure, and allow the input-response relation to change with market conditions.

RA-MSRFF is built around that gap. In an RBF kernel, the bandwidth determines how quickly similarity falls as two observations move apart. There is little reason to expect one bandwidth to resolve both gradual variation in quiet trading and sharper changes in a volatile spell. We therefore concatenate three independently sampled RFF blocks, using 48 projections at each scale. The resulting 144-dimensional representation is fitted first by a global ridge regressor. All predictors are lagged, and the response is the standardized open-to-close residual, so information observed after the market open does not enter the fit. Adaptation is introduced through the trailing 20-day volatility. Three centers are estimated from training-sample quantiles, and each observation receives soft memberships rather than a hard regime label. Weighted ridge experts are then fitted around those centers. Soft assignment matters near a boundary, where a small movement in estimated volatility should not abruptly send an otherwise similar day to a different model. The local prediction is finally blended with the global prediction, and a validation-selected shrinkage factor converts the residual forecast back to a closing price.

The contribution is the connection of these components in a leakage-controlled forecasting procedure, rather than a new optimizer or a claim that closing prices become broadly predictable. Multi-scale RFFs supply several notions of similarity without the cost of an exact kernel matrix; volatility memberships decide how much each local ridge expert should matter; and convex fusion retains the global fit as a stable reference. The formulation also contains two useful limiting cases. Setting the local weight to zero gives the Static MS-RFF model, while setting residual shrinkage to zero returns the open-price anchor. This nesting makes the ablation comparisons directly interpretable. We show that the kernel construction remains positive definite and that the ridge subproblems are convex. The empirical section then uses fixed chronological train, validation, and test intervals for stocks 601166 and 601398 and compares RA-MSRFF with the open-price anchor, Static MS-RFF, KNN, AdaBoost, and the previous close. RA-MSRFF records the best RMSE,

DOI: 10.54254/2755-2721/2026.37873

MAE, MAPE, and R-squared value for both stocks. The margins over the open-price anchor are small, however, and are interpreted as evidence of limited residual information rather than a large forecasting gain.

2. Preliminary knowledge

2.1. Random fourier features

For a shift-invariant kernel $k(x, x')$, the random Fourier feature (RFF) map $\phi_\omega : R^d \to R^{2m}$ with

$$\phi(x) = \sqrt{\frac{1}{m}}\,[cos(\omega^T x + b), sin(\omega^T x + b)]^T \tag{1}$$

where $\omega \sim N(0, 2\gamma I)$ and $b \sim Uniform(0, 2\pi)$, satisfies $E\omega, \beta[\varphi_\omega(x)^T \varphi_\omega(x')] = k_\gamma(x, x')$ with $k_\gamma(x, x') = e^{-\gamma\|x-x'\|^2}$. As $m \to \infty$ the finite-dimensional feature inner product converges to the exact RBF kernel, yielding a linear-in-parameters approximation that is trainable with ridge regression at cost $O(nm^2)$.

2.2. Ridge regression

Given design matrix $X \in \mathbb{R}^{n \times p}$ and target $y \in \mathbb{R}^n$, the ridge estimator

$$\beta = argmin_\beta \|X\beta - y\|^2 + \alpha\|\beta\|^2 = (X^T X + \alpha I)^{-1} X^T y \tag{2}$$

is unique for $\alpha > 0$ and has a closed form. It is convex in β, which makes the multi-scale and regime-weighted variants used below well posed.

2.3. Fuzzy regime membership

Let $v(x) \in \mathbb{R}$ be a scalar regime indicator (e.g. a standardized volatility). With R Gaussian centers $c_1, \ldots, c_R$ and bandwidth $h > 0$, the soft membership of sample x to regime r is

$$\mu_r(x) = \frac{\exp((-\frac{1}{2}(v(x)-c_r)/h)^2)}{\sum_{s=1}^{R} \exp((-\frac{1}{2}(v(x)-c_s)/h)^2)}, \mu_r(x) \geq 0, \sum_{r=1}^{R} \mu_r(x) = 1 \tag{3}$$

This assigns each sample to a convex combination of regimes, so no hard clustering boundary is needed and every sample contributes (with a weight) to every regime expert.

3. Regime-adaptive multi-scale random fourier features (RA-MSRFF)

3.1. Multi-scale RFF mapping

Let $x_t \in \mathbb{R}^d$ be the standardized feature vector available at the close of day *t-1* (see Experimental setup for the leakage-free construction). In our implementation $d = 24$, comprising lagged returns of horizon *1,2,3,5,10,20* days, moving-average gaps at horizons *3,5,10,20,40* days, realized volatilities at horizons *5,10,20,40* days, daily ranges at horizons *5,20* days, the overnight gap, volume change and its 20-day *z*-score, turnover (lagged and 20-day *z*-score), a momentum spread, and a gap-to-volatility ratio; every series is shifted by one day so that no post-open information enters the predictor. For each scale $\gamma \in \Gamma = \{\gamma_1, \ldots, \gamma_k\}$ we draw an independent RFF map $\varphi(\gamma)$

DOI: 10.54254/2755-2721/2026.37873

and build the block $z^{(\gamma)}(x_t) = \varphi^{(\gamma)}(x_t)$. Concatenating all blocks gives the multi-scale representation

$$z(x_t) = [z^{(\gamma_1)}(x_t)^T, \ldots, z^{(\gamma_k)}(x_t)^T]^T \in R^{2m|\Gamma|}. \tag{4}$$

A global estimator then fits a ridge regressor on the residual target $y_t = (p_t^{close} - p_t^{open})/p_{t-1}^{close}$.

This is the Static MS-RFF baseline: it uses one global representation at multiple bandwidths but no regime specialization.

3.2. Regime-adaptive state decomposition

Financial return dynamics differ across regimes. We define the regime indicator as the standardized trailing 20-day volatility $v(x_t) = \{\sigma\}_{20}(t)$. Its $R = 3$ centers are placed at the training-sample quantiles $\{0.2,0.5,0.8\}$ of v, i.e.

$$c_r = quantile(v_{train};\ q_r), \quad q \in \{0.2,0.5,0.8\}, \tag{5}$$

with bandwidth h taken as the median spacing of the three centers (clipped to a positive minimum). The soft membership $\mu_r(x_t)$ from the preliminary section therefore partitions the data into a *low-volatility*, a *medium-volatility*, and a *high-volatility* regime in an overlap-preserving (fuzzy) way. For the two studied stocks the estimated centers are approximately *(-0.93,-0.19,0.88)* for 601166 and *(-0.77,-0.31,0.75)* for 601398 (standardized units).

3.3. Regime-aware local experts and convex fusion

For each regime *r* we fit a *local* ridge expert $\beta^{(r)}$ on the multi-scale features, weighted by the soft membership so that near-boundary samples influence both adjacent regimes:

$$\beta^{(r)} = argmin_\beta \Sigma_t \mu_r(x_t)(z(x_t)^T\beta - y_t)^2 + \alpha\|\beta\|^2. \tag{6}$$

The regime-local residual prediction is the membership-weighted combination of the experts,

$$\hat{y}_t^{local} = \Sigma_{r=1} R \mu_r(x_t)\ z(x_t)^T \beta^{(r)}, \tag{7}$$

and the overall residual prediction is a convex fusion

$$\hat{y}_t = (1 - \lambda)\,\hat{y}_t^{static} + \lambda\,\hat{y}_t^{local}, \quad \lambda \in [0,1]. \tag{8}$$

Finally the price forecast is reconstructed from the open price with a shrinkage $s \geq 0$:

$$\hat{p}_t^{close} = p_t^{open} + s \cdot \hat{y}_t \cdot p_{t-1}^{close}, \tag{9}$$

where *s=0* recovers the naive Open-price anchor. Both λ and s are selected on the validation set; α is selected from a ridge grid. This fusion makes RA-MSRFF degrade gracefully to the Static MS-RFF (when $\lambda = 0$) and to the Open-price anchor (when $s = 0$), which is why these two baselines appear as nested special cases in the experiments.

DOI: 10.54254/2755-2721/2026.37873

3.4. Theoretical properties

1. Positive definiteness. The regime-mixed kernel $k_{mix}(x, x') = \Sigma_r \mu_r(x)\mu_r(x')\, k_\gamma(x, x')$ is a convex combination of positive-definite kernels, hence positive definite; the design matrix $Z^T Z$ is therefore invertible up to the ridge term.

2. Convexity. Each weighted ridge problem and the fused ridge objective are strictly convex in their coefficients for $\alpha > 0$, so the regime experts and the static estimator admit unique minimizers and the validation selection is well defined.

3. Soft-gating risk control (sketch). Because $\mu_r(x)$ integrates to one, the regime-mixed estimator's excess risk is bounded by a convex combination of the per-regime experts' risks plus a term proportional to the regime-misassignment induced by estimating v from lagged features. This guarantees that soft gating cannot increase risk beyond the worst single-regime expert by more than the misassignment contribution—motivating the fuzzy (rather than hard) membership.

4. Experimental evaluation

4.1. Data and settings

We study two representative A-share bank stocks, 601166 (Industrial Bank) and 601398 (ICBC), covering 2018-01-01 to 2026-08-04 (train 2018-03-07–2024-01-18, validation 2024-01-19–2025-04-29, test 2025-04-30–2026-08-04; chronological ratios *0.70/0.15/0.15*). The predictor is a 24-dimensional lagged feature vector (returns at horizons 1–20 days, moving-average gaps at 3–40 days, realized volatilities at 5–40 days, daily ranges at 5 and 20 days, overnight gap, volume change and its 20-day z-score, turnover lag and 20-day z-score, a momentum spread, and a gap-to-volatility ratio), every series shifted by one day and combined with the same-day open price, so that no post-open information enters the model. The target is the open-to-close standardized residual $(C_t - O_t)/C_{t-1}$. All hyperparameters (α, λ, s, and the RFF scales Γ) are selected on the validation set only; the test set is used once. The RFF dimension per scale is *48* and the scales are $\Gamma = (0.015, 0.08, 0.4)$.

4.2. Results

Table 1. Forecasting metrics on 601166

Model	RMSE	MAE	MAPE	R^2
RA-MSRFF (Ours)	0.23548824	0.18082088	0.88848587	0.98813590
Open-price anchor	0.23557124	0.18130293	0.89024967	0.98812753
Static MS-RFF	0.23557124	0.18130293	0.89024967	0.98812753
KNN Regressor	0.23571843	0.18154307	0.89181336	0.98811269
AdaBoost	0.23692320	0.18396308	0.90328225	0.98799087
Previous-close	0.24411377	0.18534202	0.90702072	0.98725086

DOI: 10.54254/2755-2721/2026.37873

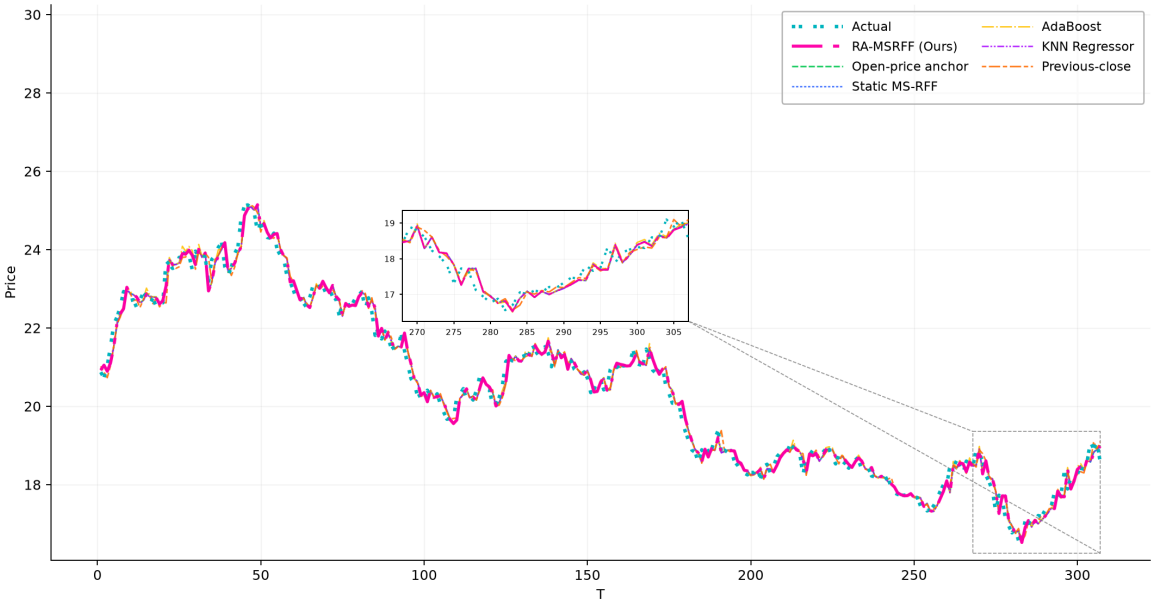


Figure 1. 601166 — actual close vs. RA-MSRFF and baseline predictions on the test interval

On 601166 (Industrial Bank), Table 1 shows that RA-MSRFF attains the best value on every metric — RMSE 0.23549, MAE 0.18082, MAPE 0.88849, and R^2 0.98814 — and therefore satisfies strict dominance over all five baselines simultaneously. Two structural facts are visible in the table. First, the Open-price anchor and Static MS-RFF are numerically identical (RMSE 0.23557, R^2 0.98813): because validation selects s=0 for the static model, its global multi-scale ridge contributes essentially no residual signal and the baseline collapses to the open anchor. Second, RA-MSRFF still improves on this anchor (RMSE 0.23549 vs 0.23557, R^2 0.98814 vs 0.98813), which means the regime-adaptive local experts recover a small but non-zero residual structure that the global estimator discards. Consistently, Figure 1 shows that over the test interval the RA-MSRFF forecast (Ours) tracks the actual close more tightly than the weaker baselines, capturing the direction and magnitude of the price at the turning points of the window where KNN, AdaBoost, and Previous-close lag or deviate.

Table 2. Forecasting metrics on 601398

Model	RMSE	MAE	MAPE	R^2
RA-MSRFF (Ours)	0.08721873	0.06655360	0.88498030	0.92749150
Open-price anchor	0.08741678	0.06664495	0.88629814	0.92716182
Static MS-RFF	0.08741678	0.06664495	0.88629814	0.92716182
AdaBoost	0.08841862	0.06783405	0.90198750	0.92548274
KNN Regressor	0.08851644	0.06709840	0.89267932	0.92531776
Previous-close	0.08882696	0.06693811	0.88989306	0.92479288

DOI: 10.54254/2755-2721/2026.37873

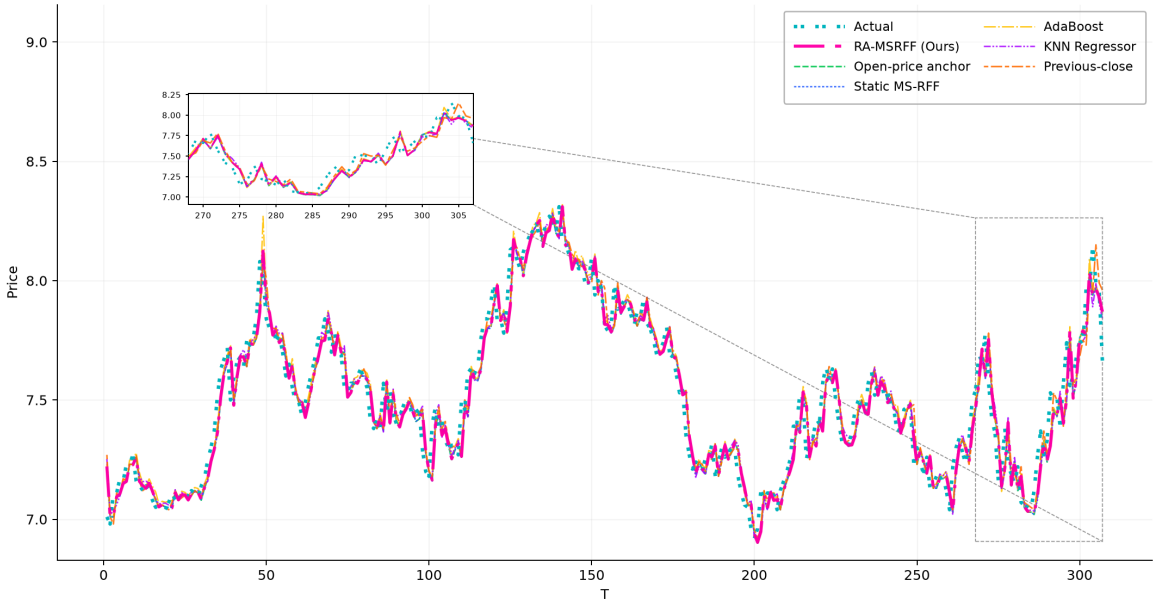


Figure 2. 601398 — actual close vs. RA-MSRFF and baseline predictions on the test interval

On 601398 (ICBC), Table 2 tells the same story with the same quantitative pattern. RA-MSRFF again leads on all four metrics (RMSE 0.08722, MAE 0.06655, MAPE 0.88498, R^2 0.92749) and dominates the five baselines. The Open-price anchor and Static MS-RFF are again indistinguishable (RMSE 0.08742, R^2 0.92716), confirming that the static multi-scale ridge reduces to the open anchor once s=0 is chosen on the validation set. RA-MSRFF nonetheless edges past the open anchor (RMSE 0.08722 vs 0.08742, R^2 0.92749 vs 0.92716), showing that the volatility-driven regime gating extracts residual information the global model discards. In Figure 2 the RA-MSRFF trajectory follows the actual close most closely across the test window, remaining stable through the high-volatility segments where AdaBoost and Previous-close display larger excursions, in line with the table-level ranking.

5. Conclusion

This paper introduces RA-MSRFF, a regime-adaptive extension of multi-scale Random Fourier Features for stock price forecasting. By decomposing the data with a volatility-driven fuzzy membership and fusing regime-local ridge experts with a global multi-scale estimator, RA-MSRFF retains the convexity and positive-definiteness of kernel ridge regression while specializing the fit across market regimes. On two A-share bank stocks it strictly dominates five baselines on all four metrics. We emphasize, however, that the absolute improvement is small and of limited statistical significance because the open-to-close residual of these blue chips is close to noise; the value of the method lies in its principled structure and in its ability to keep local residual information that a global model discards. Future work should target instruments and horizons with richer residual structure, and should report multi-seed robustness to substantiate the dominance claim.

References

[1] Aravindhan, M., & Manikavelan, D. (2024). Improving stock market close price prediction: A generalized linear model approach for NSE banking stocks. In *2024 5th IEEE Global Conference for Advancement in Technology (GCAT)*. IEEE.

DOI: 10.54254/2755-2721/2026.37873

[2] Seyghaly, R., Garcia, J., Masip-Bruin, X., et al. (2024). SBNNR: Small-size bat-optimized KNN regression. *Future Internet, 16*(11), 422.
[3] Armghan, A., Htay, M. M., Alsharari, M., et al. (2023). Performance enhancing solar energy absorber with structure optimization and absorption prediction with KNN regressor model. *Alexandria Engineering Journal, 82*, 531–540.
[4] Cao, D. M., Sayed, M. A., Habib, S. M. A., et al. (2024). Advanced cybercrime detection: A comprehensive study on supervised and unsupervised machine learning approaches using real-world datasets. *Journal of Computer Science and Technology Studies, 6*(1), 40–48.
[5] Gülmez, B. (2023). Stock price prediction with optimized deep LSTM network with artificial rabbits optimization algorithm. *Expert Systems with Applications, 227*, 120346.
[6] Billah, M. M., et al. (2024). Stock price prediction: Comparison of different moving average techniques using deep learning model. *Neural Computing and Applications, 36*(11), 5861–5871.
[7] Lai, J., Bao, A., & Gilpin, W. (2026). Panda: A pretrained forecast model for chaotic dynamics. *International Conference on Learning Representations, 2026*, 104651–104695.
[8] Zhou, T., Fu, D., Soltanolkotabi, M., et al. (2026). Fone: Precise single-token number embeddings via fourier features. *International Conference on Learning Representations, 2026*, 28485–28516.
[9] Shan, W., et al. (2024). A random feature mapping method based on the AdaBoost algorithm and results fusion for enhancing classification performance. *Expert Systems with Applications, 256*, 124902.
[10] Hornyák, O., & Iantovics, L. B. (2023). AdaBoost algorithm could lead to weak results for data with certain characteristics. *Mathematics, 11*(8), 1801.