

Learning Adaptive Multi-Scale Memory Kernels in Fractional Stochastic Volterra Models: An AI-Assisted Framework for Financial Volatility

Gauri N. Phad^{1,*} and V. R. Nikam¹

Department of Mathematics and Research, L. V. H. ASC College (Autonomous), Affiliated to Savitribai Phule Pune University, Panchavati, Nashik, Maharashtra, India

Corresponding author: gauri.phdwork@gmail.com

Abstract

We develop an interpretable volatility-forecasting framework in which positive Elastic Net selects a finite power-law memory dictionary and Gaussian-mixture probabilities combine regime-specific variance forecasts. A conditional product-integration theorem gives a grid-error bound on every fixed finite horizon by an exponentially weighted argument. A reconstructed evaluation of NIFTY 50 daily OHLC data fits fully observed labels through 2023, purges overlapping validation targets and compares nine models at identical 2024–2026 origins. Soft adaptation lowers twenty-day QLIKE relative to the static dictionary in all three periods, but the paired advantages are inconclusive; it increases five-day loss in 2024. Conditional block resampling exposes uncertainty in memory composition. In a retained NIFTY-to-SPY transfer experiment, a source-trained constant has lower average errors than the fractional forecast. The contribution is a transparent construction linking fractional memory features, sparse learning and probabilistic adaptation with a stated numerical guarantee and auditable evidence about the usefulness of added complexity. Regression importance describes predictive scales rather than calibrated stochastic-kernel parameters. All empirical evaluations are retrospective.

Keywords: fractional stochastic Volterra equation; power-law memory; Elastic Net; regime adaptation; volatility forecasting; Garman–Klass variance.

1 Introduction

Volatility clustering and persistent dependence motivate models whose present state depends on past observations. Classical diffusion pricing models [1, 2] remain useful reference points, while long-memory, fractional and rough-volatility constructions represent non-local dependence more explicitly [3, 4, 7, 8]. A practical difficulty is deciding which memory scales are useful at a particular forecast horizon without adding unjustified complexity.

We address this choice with a finite fractional dictionary, non-negative sparse regression and probabilistic regime experts. The continuous-time Volterra equation provides a mathematical object for kernel interpretation and numerical analysis. Its power-law shapes also define causal features of observable daily range variance. The empirical regression is a separate reduced-form forecasting construction, and its coefficients are not estimates of every parameter in the stochastic equation.

The contribution combines explicit feature normalization and coefficient back-transformation, variance-scale soft gating, a conditional finite-horizon error proof, and a timing-audited comparison with simple and dynamic baselines. The experiments ask whether useful scales vary with horizon, whether regime adaptation adds predictive value, and whether transfer accuracy exceeds source-level matching. These questions support a specific assessment of memory complexity rather than an assertion of universal forecasting superiority.

2 Related Work and Positioning

Fractional calculus supplies hereditary power-law operators [5, 6]. For $\gamma > 0$, the Riemann–Liouville integral is

$$I^\gamma \phi() = \frac{1}{\Gamma(\gamma)} \int_0^t ()^{\gamma-1} \phi() \, ds. \quad (1)$$

Long-memory and rough-volatility models [3, 4, 8] specify structured dependence. Path-dependent weighting [21], deep stochastic-model calibration [25] and alternative Volterra kernels [30] motivate flexible memory representations. GARCH, EGARCH, GJR-GARCH, FIGARCH and HAR capture persistence and asymmetry through econometric dynamics [9–13]. Regularized forecasting studies report horizon-dependent predictor relevance [20, 26, 27], while broader comparisons give mixed rankings for linear and machine-learning methods [28, 29, 33]. Regime-switching AR/HAR models provide a related route to state dependence [32].

Here regimes alter the composition of an observable fractional dictionary through separate sparse experts. The individual tools are established; the contribution is their explicit mathematical connection and audited evaluation. Persistence of a measured proxy does not identify a unique latent roughness law [24].

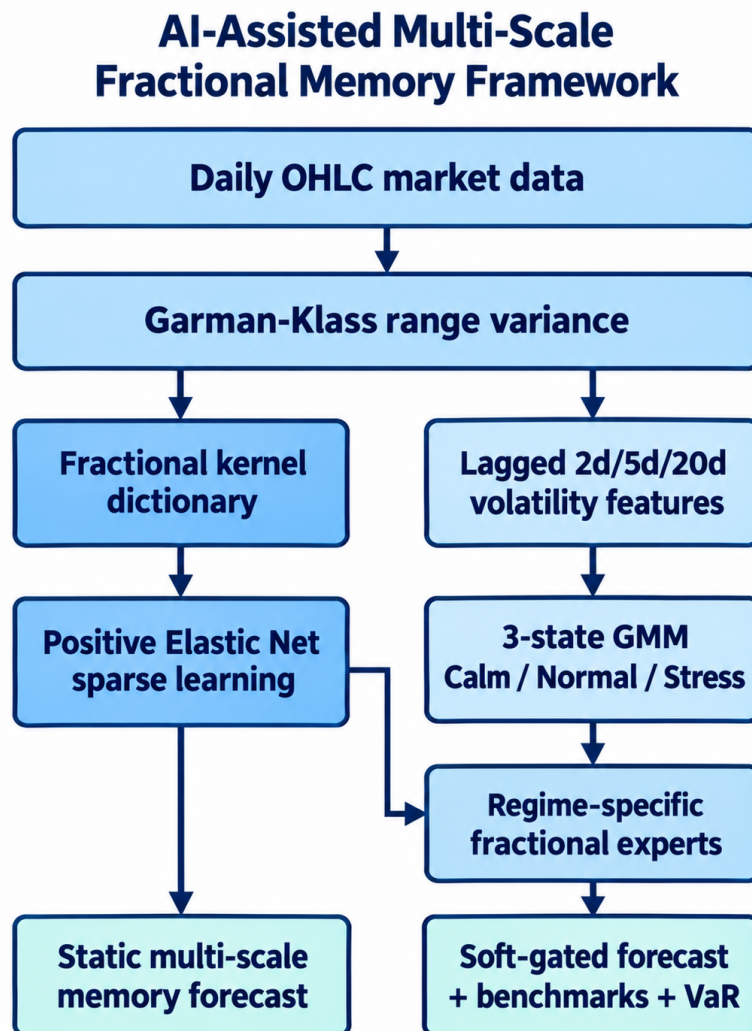


Figure 1. Fractional-memory learning and probabilistic adaptation. VaR is an exploratory downstream diagnostic.

3 Fractional Stochastic Volterra Framework

3.1 Continuous-time model and admissible kernels

On a filtered probability space satisfying the usual conditions, let $X_0 \in L^2()$ be F_0 -measurable. For fixed finite coefficients we consider

$$X() = X_0 + I^\alpha[]() + \sum_{m=1}^M \lambda_m I^{\alpha+\beta_m}[]() + \Xi(). \quad (2)$$

Here $0 < \alpha < 1$, $0 < \beta_m < 1$, $\lambda_m \geq 0$, and Ξ is a specified adapted additive process with $\Xi() = 0$. The numerical theorem assumes a unique square-integrable solution and its required regularity. Non-negative coefficients alone neither establish existence nor make X a positive variance process. The order α belongs to the Volterra backbone; the β_m index hereditary channels.

$$K_\theta() = \sum_{m=1}^M \lambda_m \frac{u^{\beta_m-1}}{\Gamma()}, \quad u > 0. \quad (3)$$

$$()() = \sum_{m=1}^M \lambda_m I^{\beta_m}V(). \quad (4)$$

Smaller β_m concentrate relatively more weight near the present; larger values decay more slowly. The finite mixture is locally integrable because

$$\int_0^T K_\theta() du = \sum_m \frac{\lambda_m T^{\beta_m}}{\Gamma()} < \infty.$$

It is completely monotone: $()^n K_\theta^0() \geq 0$ for every integer $n \geq 0$. This follows term by term from the derivative signs of u^{β_m-1} . If every G_m equals a common G , the fractional-integral semigroup identity yields $\sum_m \lambda_m I^{\alpha+\beta_m}G = I^\alpha()$. Distinct G_m retain separate channels and do not admit that reduction.

3.2 Deterministic memory and stochastic integration

Local integrability is distinct from the condition for an Itô convolution. For Brownian motion W and predictable a , the process $\Xi() = \int_0^t K_\theta()a() dW_s$ requires $\int_0^t ||^2 E||^2 ds < \infty$. A sufficient condition is $K_\theta \in L^2()$ with uniformly bounded second moments of a . For a nonzero non-negative mixture (3), square integrability holds exactly when every active exponent exceeds $1/2$; the smallest active exponent controls the singularity. The forecasting dictionary includes smaller exponents because it weights observed variance rather than Brownian increments.

Consequently, the empirical exercise does not calibrate (2), its order α , its physical coefficients λ_m , or a latent Hurst parameter. The shared power-law family connects model interpretation to observable features, while the mathematical and empirical claims retain their separate assumptions.

3.3 Causal features and sparse scale learning

Let V_t be daily Garman–Klass variance observed after the close at origin t . For $L = 252$ and the dictionary $\beta \in \{0.1, 0.2, \dots, 0.9\}$, define

$$M_\beta() = \sum_{j=1}^L w_j^0 V_{t-j+1}, \quad w_j^0 = \frac{j^{\beta-1}}{\sum_{\ell=1}^L \ell^{\beta-1}}. \quad (5)$$

The normalized weights isolate memory shape and include the observed origin day. They are discrete forecasting weights, distinct from the product-integration weights of Section 5. For $h \in \{5, 20\}$, the future target is

$$Y_{t,h} = \log(), \quad \varepsilon = 10^{-12}. \quad (6)$$

Set $Z_{t,m} = []/s_m$, where μ_m and $s_m > 0$ use the current training fold only. Positive Elastic Net [18] estimates

$$() = \arg \min_{b_0 \in \mathbb{R}, b \geq 0} \left\{ \frac{1}{2N} \sum_{t=1}^N ()^2 + \tau [] \right\}. \quad (7)$$

The intercept is unrestricted. The penalty $\tau > 0$ and mixture $\eta_E \in []$ are distinct from the fractional parameters. Back-transform coefficients by $\hat{c}_m = \hat{b}_m/s_m$. When their sum is positive, report

$$\omega_m = \frac{\hat{c}_m}{\sum_{\ell: \hat{c}_\ell > 0} \hat{c}_\ell}, \quad \sum_m \omega_m = 1. \quad (8)$$

$$\beta_{\text{eff}} = \sum_m \omega_m \beta_m. \quad (9)$$

These shares normalize original log-feature coefficients, not standardized coefficients. The effective exponent summarizes predictive composition; a finite mixture and a finite feature window do not identify a single physical memory law. Both summaries are undefined for an intercept-only fit.

$$\hat{V}_{t,h} = \exp() - \varepsilon. \quad (10)$$

The residual variance $\hat{s}_h^2$ uses training residuals and denominator $N-1$. The Gaussian factor is exact for centered log residuals with constant conditional variance and otherwise is a plug-in approximation. Development checks favour omitting it, but the main documented forecasts retain it. Equivalently,

$$\hat{V}_{t,h} + \varepsilon = A_h \prod_m ()^{\hat{c}_m}, \quad A_h = \exp().$$

For fixed parameters, non-negative coefficients make the forecast non-decreasing in each memory feature. This multiplicative form also distinguishes predictive shares from additive kernel coefficients. Forecasts and loss arguments are floored at ε before evaluation.

4 Probabilistic Regime Adaptation

A three-component Gaussian mixture uses trailing variance information available at each origin.

Write $V_{t,k} = k^{-1} \sum_{j=0}^{k-1} V_{t-j}$ and define

$$R_t = []^T. \quad (11)$$

Protect logarithms in (11) with ϵ and standardize R_t with training moments to obtain z_t . Order components by training-period twenty-day volatility and label them Calm, Normal and Stress. These are fitted clusters, without an independently observed state or a fitted Markov transition law. Posterior probabilities are

$$\pi_r() = \frac{a_r \varphi_3()}{\sum_{\ell=1}^3 a_\ell \varphi_3()}, \quad r = 1, 2, 3,$$

where $a_r \geq 0$, $\sum_r a_r = 1$, and φ_3 is a Gaussian density with positive-definite covariance. Each state fits a separate positive Elastic-Net expert using the same dictionary and the horizon-specific penalty selected for the static model.

$$\hat{V}_{t,h}^{\text{hard}} = \hat{V}_{t,h}^{()} \quad r^* = \operatorname{argmax}_r \pi_r(). \quad (12)$$

$$\hat{V}_{t,h}^{\text{soft}} = \sum_{r=1}^3 \pi_r() \hat{V}_{t,h}^{()}. \quad (13)$$

Soft gating averages retransformed variances, so it propagates classification uncertainty and lies between the minimum and maximum expert forecasts. Averaging log forecasts before exponentiation produces a different model. Posterior-weighted coefficient shares likewise do not generally reconstruct this nonlinear variance mixture.

4.1 Training and model selection

The reconstructed static search retains 17 logarithmically spaced τ values from 10^{-5} to 10^{-1} and six mixtures $\{0.1, 0.3, 0.5, 0.8, 0.95, 1\}$. Mean QLIKE over five expanding, purged folds selects $() = ()$ at five days and $()$ at twenty days. A single-power comparator selects its exponent and penalty on the same folds; it selects $\beta = 0.1$ at both horizons.

The GMM uses full covariances, seed 42, ten initializations, covariance regularization 10^{-6} , tolerance 10^{-5} and at most 1,000 iterations. Experts require at least 12 training observations and receive no additional tuning on evaluation outcomes. Elastic Net uses non-negative coefficients, at most 300,000 iterations and tolerance 10^{-7} . Feature standardization, mixture fitting and residual corrections use retained training observations only. Final fits use fully observed labels ending by 29 December 2023 and remain frozen during evaluation.

5 Numerical Approximation and Convergence

Singular-kernel Volterra discretization has driver-dependent error behaviour [22, 23, 31, 34]. For the fixed finite model (2), let $t_n = n\delta$, $\delta = T/N$. Forecast horizon h remains distinct from mesh width δ . Define

$$a_k^{(\gamma)} = \frac{k^{\gamma} - (\gamma)}{\Gamma(\gamma)}, \quad k \geq 1. \quad (14)$$

$$\begin{aligned} X_n = & X_0 + \delta^{\alpha} \sum_{j=0}^{n-1} a_{n-j}^{(\gamma)} F() \\ & + \sum_{m=1}^M \lambda_m \delta^{\alpha+\beta_m} \sum_{j=0}^{n-1} a_{n-j}^{(\gamma)} G_m() + \Xi_n. \end{aligned} \quad (15)$$

5.1 Conditional error on a fixed finite horizon

Theorem 1. Assume that (2) has a unique square-integrable solution on $[0, T]$, with the same initial value in (15) and $\Xi(0) = \Xi_0 = 0$. Suppose F and G_m are globally Lipschitz in the state, with constants L_F, L_m . Assume the composed integrands $F()$ and $G_m()$ are uniformly η -Hölder continuous in L^2 on $[0, T]$, with $\eta \in (0, 1]$ and constants H_F, H_m . Choose a fixed $\xi > 0$ such that

$$q_{\xi} = L_F \xi^{-\alpha} + \sum_m \|L_m \xi^{-\alpha}\| < 1. \quad (16)$$

Such a choice exists because all orders are positive and the mixture is finite. Suppose, for some $\rho > 0$ and mesh-independent C_{Ξ} , that

$$\max_{0 \leq n \leq N} \|\Xi_n - \Xi_n\|_{L^2} \leq C_{\Xi} \delta^{\rho}. \quad (17)$$

Then the maximum grid error obeys

$$\max_{0 \leq n \leq N} \|X_n - X_n\|_{L^2} \leq \frac{e^{\xi T}}{1 - q_{\xi}}(), \quad (18)$$

where $C_0 = H_F T^{\alpha} / \Gamma(\gamma) + \sum_m \|H_m T^{\alpha+\beta_m} / \Gamma(\gamma)\|$. Therefore the grid error is $O(\delta^{\rho})$ on every fixed finite horizon. Its constant depends on T and the kernel, not on mesh width.

The exponential weight removes a smallness restriction on the corresponding unweighted finite-horizon Lipschitz bound. This result does not prove existence, a universal fractional-Brownian-motion convergence order, or uniform stability as $T \rightarrow \infty$. Assumption (17) must be justified for the chosen stochastic driver, integration convention and quadrature.

5.2 Proof and manufactured verification

Proof. For an L^2 -valued η -Hölder integrand ϕ , integrate the kernel exactly on each mesh interval and replace $\phi()$ by its left-endpoint value. Minkowski's inequality gives

$$\|\phi\|_{L^2} \leq H_\phi \delta^\eta \frac{t_n^\gamma}{\Gamma()}. \quad (14)$$

The combined deterministic consistency residual is at most $C_0 \delta^\eta$. Let $e_n = \|X() - X_n\|_{L^2}$ and $E_{\xi,N} = \max_n e_n e^{-\xi t_n}$. For $s \in []$, we have $e_j \leq e^{\xi t} E_{\xi,N}$ and $e^{-\xi()} \leq e^{-\xi()}$. For each $\gamma > 0$, the weighted stability integral is bounded by

$$\frac{1}{\Gamma()} \int_0^{t_n} e^{-\xi u} u^{\gamma-1} du \leq \xi^{-\gamma}.$$

Subtracting (15) from (2), applying the Lipschitz bounds and assumption (17), and taking the weighted maximum therefore yields $E_{\xi,N} \leq C_0 \delta^\eta + C_\Xi \delta^\rho + q_\xi E_{\xi,N}$. Divide by $1 - q_\xi$ and use $\max_n e_n \leq e^{\xi T} E_{\xi,N}$ to obtain (18). $\square$

For deterministic verification set $X_0 = 0$, $\Xi = 0$ and consider

$$X() = I^\alpha f() + \sum_{m=1}^M \lambda_m I^{\alpha+\beta_m} X(). \quad (19)$$

The manufactured solution $X() = t^\gamma$ has forcing

$$f() = \frac{\Gamma()}{\Gamma()} t^{\gamma-\alpha} - \sum_m \lambda_m \frac{\Gamma()}{\Gamma()} t^{\gamma+\beta_m}. \quad (20)$$

The retained configuration uses $\alpha = 0.7$, $\beta = ()$ and illustrative non-negative coefficients scaled to the relative twenty-day importance proportions. These coefficients are numerical demonstration parameters, not stochastic-equation estimates. Across meshes $N = 25, \dots, 1600$, the observed maximum-norm orders are summarized in Table 1.

Table 1. Deterministic manufactured verification with zero stochastic convolution.

Solution	Observed order	Uniform integrand bound
Smooth: $X() = t^2$	1.002	$\eta = 1$
Weak: $X() = t^{0.8}$	0.800	$\eta = 0.1$

For the weak example, forcing starts with $t^{\gamma-\alpha} = t^{0.1}$. The theorem therefore supplies only the conservative uniform Hölder exponent 0.1; the observed 0.800 order needs sharper endpoint analysis to be proved. These deterministic examples do not verify the stochastic rate in (17).

6 Data and Evaluation Protocol

The annual NIFTY 50 OHLC files contain 1,413 observations: 1,239 for 2021–2025 and 174 for 2026, with the last observed trading day on 15 September 2026. There are no duplicate dates, and the positive-price OHLC consistency checks pass. The Garman–Klass proxy [17] is

$$v_t^{\text{GK}} = \frac{1}{2} [r_t^2 - (r_t^-)^2 - (r_t^+)^2] \quad (21)$$

This is a daily range-based variance proxy, not intraday realized variance. Pre-2024 lag-one autocorrelation is approximately 0.359, with strong serial dependence at twenty lags. Such diagnostics motivate memory features but do not identify latent roughness [24].

Training uses only forward labels ending by 29 December 2023, giving 486 five-day and 471 twenty-day observations. At an after-close origin t , a training origin u is eligible only if $u + h \leq t$. Five expanding validation folds omit $h-1$ adjacent origins, with gaps of 4 and 19. All training moments and model components use eligible observations; final parameters are frozen. Subsequent observations update causal features and EWMA only. Ten fold-boundary audits and all exported prediction cutoffs pass; perturbing future observations does not alter features at the tested origins.

Table 2. Evaluation blocks by forecast-origin year. Forward targets are fully observed.

Origin year	5-day N	20-day N	Target timing
2024	248	248	Targets may end in January 2025
2025	249	249	Targets may end in January 2026
2026	169	154	Last origins: 7 September / 17 August

Nine models share targets and origins: static dictionary, selected single power, hard and soft experts, direct log-OLS HAR-GK, EWMA-GK with decay 0.94, Historical-20, current-GK persistence and a training-label mean. HAR uses daily, five-day and twenty-day trailing log GK averages with the same Gaussian correction. The original GARCH-family fits were not regenerated and are excluded from the reconstructed contest. Historical forecasts and tests remain archived.

QLIKE is $L(\hat{v}) = \hat{v}/v - \log(\hat{v}) - 1$, with arguments floored at ε . Its proxy-robust ranking property needs additional assumptions [19]; using the loss does not establish proxy unbiasedness. RMSE, MAE and correlation compare annualized volatility $\sqrt{252}v$. For differences $d_t = L_t^A - L_t^B$,

$c_t = d_t - d$, inference [14] uses

$$\hat{S}_\ell = \frac{1}{N} \sum_t c_t^2 + 2 \sum_{k=1}^{\ell} \left(\frac{1}{N} \sum_{t=k+1}^N c_t c_{t-k} \right), \quad \text{DM} = \frac{d}{\sqrt{\hat{S}_\ell / N}}.$$

Two-sided normal-reference values use bandwidth $\ell = h-1$, with $2h-1$ sensitivity. Eight comparisons against Static form a separate retrospective Holm family [35] for each horizon/year. Paired circular blocks of lengths h and $2h$ use 2,000 resamples; their 95% intervals are nominal, not multiplicity-adjusted. Historical inspection and architecture choices make all blocks retrospective, including 2026.

7 Reconstructed Forecasting Results

Table 3. Origin-purged QLIKE on matched targets. Lower is better; sample sizes are in Table 2.

Model	2024 5d	2024 20d	2025 5d	2025 20d	2026 5d	2026 20d
Static	0.47794	0.35673	0.07764	0.05324	0.22409	0.24569
Single power	0.47564	0.32067	0.07806	0.05394	0.22333	0.21999
Hard regime	0.52063	0.28825	0.08165	0.04434	0.23199	0.18379
Soft regime	0.52631	0.26678	0.07807	0.04539	0.23099	0.18737
HAR-GK	0.37209	0.22726	0.08334	0.05760	0.16450	0.16179
EWMA-GK	0.41545	0.26720	0.07584	0.05497	0.20577	0.22041
Historical-20	0.42160	0.28808	0.07649	0.05865	0.22294	0.24703
Persistence	1.14578	1.26382	0.37271	0.40531	0.49755	0.51628
Training mean	0.30978	0.16391	0.21625	0.16490	0.24174	0.15142

In 2024, hard and soft adaptation increase five-day QLIKE by 0.042685 and 0.048372, with Holm-adjusted $p = 0.003360$ and 0.012025 . At twenty days, soft adaptation lowers mean loss by 0.089947, but adjusted $p = 0.932558$ and the nominal block interval $[\]$ are inconclusive. The corresponding adjusted values in 2025 and 2026 are 1.000000 and 0.491161; their intervals also include zero. Wider bandwidths and blocks preserve that interpretation.

A training mean has the lowest observed twenty-day QLIKE in 2024 and 2026, while regime experts have the lowest values in 2025. The ranking therefore depends on period and comparator. Because origin-year assignment and reconstructed fitting details differ from the earlier implementation, score changes cannot be attributed to purging alone.

Table 4. NIFTY twenty-day performance in 2026, $N = 154$. Volatility is an annualized fraction.

2026 model	QLIKE	Vol. RMSE	Vol. MAE	Correlation
Static	0.24569	0.03184	0.02558	0.14073
Single power	0.21999	0.03074	0.02489	0.20406
Hard regime	0.18379	0.02813	0.02170	0.39704
Soft regime	0.18737	0.02842	0.02214	0.37484
HAR-GK	0.16179	0.02807	0.02345	0.38376
EWMA-GK	0.22041	0.03104	0.02681	0.29187
Historical-20	0.24703	0.03266	0.02638	0.32072
Persistence	0.51628	0.04643	0.03507	0.35801
Training mean	0.15142	0.02900	0.02431	n/a

The 2026 hard and soft correlations are 0.3970 and 0.3748, versus 0.1407 for Static. HAR has lower QLIKE and slightly lower RMSE than the adaptive forecasts; hard gating has the lowest MAE. The constant has lowest QLIKE but undefined correlation. Mean-level accuracy and dynamic tracking therefore require separate assessment.

7.1 Memory Stability and Development Sensitivity

At five days, Static assigns 98.28% and 1.72% of original-scale importance to $\beta = 0.1$ and 0.2, with $\beta_{\text{eff}} = 0.1017$. At twenty days the shares on 0.1,0.2,0.3 are 42.38%, 13.78% and 43.84%, with effective exponent 0.2015. Nearby features are strongly correlated: pairwise ranges are 0.8730–0.9983 and 0.8682–0.9981, with standardized condition numbers approximately 4.62×10^7 and 4.95×10^7 .

Table 5. Effective-exponent stability from 200 conditional training-block resamples of length 20.

Horizon / state	N	Fitted exponent	Median exponent	2.5–97.5% range	Zero fits
5d Static	486	0.1017	0.1037	[0.1000, 0.3151]	0
5d Calm	209	0.1000	0.1195	[0.1000, 0.9000]	4
5d Normal	179	0.1483	0.1909	[0.1000, 0.7826]	0
5d Stress	98	0.1559	0.1706	[0.1000, 0.4854]	0
20d Static	471	0.2015	0.2559	[0.1000, 0.6914]	0
20d Calm	196	0.9000	0.9000	[0.1000, 0.9000]	1
20d Normal	189	0.1112	0.1633	[0.1000, 0.9000]	18
20d Stress	86	0.1082	0.2000	[0.1000, 0.6355]	2

Static resamples condition on selected penalties; expert resamples also fix the trained classifier. Full chronological blocks are drawn before state selection. All 200 fits per row are numerically valid; exponent percentiles use nonzero fits only. Broad ranges and intercept-only experts prevent structural identification. Seeds 0–4 versus 42 give adjusted Rand ranges 0.9729–1.0000 and 0.9894–1.0000, with maximum relative soft-forecast changes 9.06% and 2.70%. Classification stability does not imply coefficient stability.

Table 6. Development-only purged validation loss with fixed selected Static penalties.

Development check	5-day QLIKE	20-day QLIKE	Cohort N: 5d / 20d
Gaussian retransformation	0.14780	0.13787	486 / 471
No retransformation	0.14457	0.12489	486 / 471
2-state soft gating	0.15706	0.14908	486 / 471
3-state soft gating	n/a	n/a	486 / 471
Dense dictionary	0.14780	0.13787	486 / 471
Coarse dictionary	0.14780	0.13787	486 / 471
126-day lookback	0.15433	0.13796	234 / 219
252-day lookback	0.14806	0.31420	234 / 219
504-day lookback	0.15934	0.12499	234 / 219

The coarse dictionary is {0.1,0.3,0.5,0.7,0.9}. Lookback rows use the same smaller late-development cohorts and the 252-day penalties, so they are conditional checks, not separately optimized contests. Three-state gating lacks a complete five-fold score because one early state has 11 observations, below the required 12; no partial-fold average is substituted. Removing the Gaussian factor lowers development loss, but the documented main forecast remains unchanged. These conditional diagnostics do not include full architecture-selection uncertainty and do not turn retrospective evaluation into prospective confirmation.

8 Cross-Market Transfer and Exploratory Risk

The retained transfer freezes the original NIFTY-trained experts and applies the same 252-day feature construction to SPY OHLC observations from Financial Modeling Prep. Its 148 origins run from 5 January to 6 August 2026, with twenty-day target ends through 3 September. No SPY-specific expert refit is reported. The added constant is the mean of 471 fully observed pre-2024 NIFTY target labels, equivalent to annualized volatility of 10.783%; no SPY outcome is used to fit it.

Table 7. Retained SPY twenty-day forecasts and the source-only constant, $N = 148$.

SPY model	QLIKE	Vol. RMSE	Vol. MAE	Correlation
Zero-shot fractional	0.19397	0.02952	0.02258	-0.18122
HAR-GK	0.16487	0.03655	0.03065	-0.03874
EWMA-GK	0.19034	0.03158	0.02743	-0.31263
Historical-20	0.26426	0.03519	0.02981	-0.16954
Source-training mean	0.08926	0.02268	0.01956	n/a

The constant improves average QLIKE, RMSE and MAE on the fractional forecast, while all four varying forecasts have negative correlations. Fractional-minus-constant QLIKE is 0.104705, with nominal $p = 0.169922$, four-comparison Holm $p = 0.679689$ and nominal block interval $[\cdot]$. Thus transfer superiority remains unconfirmed. The four-comparison family and retained earlier statistics use explicitly distinguished HAC conventions in the archive.

8.1 Risk-scale mapping and saved backtests

For the exploratory one-day risk scale, the saved experiment maps the twenty-day expected GK average using a fixed NIFTY 2021–2023 variance bridge:

$$\kappa = \frac{\text{mean}(\cdot)}{\text{mean}(\cdot)} \approx 1.6502. \quad (22)$$

$$\hat{\sigma}_{t+1} = \sqrt{\kappa V_{t,20}^{\text{GK}}}. \quad (23)$$

For left-tail probability α , zero conditional return mean and standard-normal distribution Φ , the return threshold is $\hat{q}_{\alpha,t+1} = \hat{\sigma}_{t+1} \Phi^{-1}(\cdot)$; loss VaR is $-\hat{q}_{\alpha,t+1}$. An exception occurs if the next close-to-close log return is below the threshold. Student-t robustness uses $\hat{\sigma}_{t+1} \sqrt{v} t_v^{-1}(\cdot)$, with training-estimated $v \approx 5.39 > 2$. The multiplier ensures unit variance. Mean quantile loss is $\langle \cdot \rangle$.

Table 8. Saved Gaussian 95% SPY VaR diagnostics [15, 16], $N = 148$.

Model	Exceptions	Rate	Kupiec p	Indep. p	CC p
Fractional	8	5.41%	0.823	0.430	0.714
HAR-GK	2	1.35%	0.016	0.814	0.054
EWMA-GK	7	4.73%	0.879	0.316	0.598
Historical-20	8	5.41%	0.823	0.337	0.615

At 99%, the fractional forecast has two exceptions, with coverage, independence and conditional-coverage p-values 0.683, 0.814 and 0.895. Few tail observations and the smoothed twenty-day input limit interpretation: non-rejection does not establish calibration. The constant baseline has not been assessed in these saved risk tests. A genuine one-day target and a longer risk sample remain necessary.

9 Discussion and Conclusion

A normalized fractional dictionary makes horizon and regime dependence interpretable. Back-transformed non-negative coefficients summarize predictive scale importance; soft gating preserves a convex mixture of expert variances. The construction specifies which quantities are learned and which remain assumptions of the stochastic model.

The conditional theorem supplies an explicit finite-horizon discretization guarantee. Exponential weighting removes a smallness restriction on the unweighted contraction factor, while the proof separates integrand regularity from stochastic-convolution error. Deterministic numerical orders therefore do not supply an unverified stochastic-driver rate.

The corrected evidence supports selective complexity. At five days, Static is nearly single-scale and regime adaptation is harmful in 2024. At twenty days, soft adaptation lowers mean QLIKE relative to Static in all three periods, but paired inference is inconclusive. In 2026, adaptive forecasts track variation better than Static, while HAR and the training mean improve different error criteria. The SPY constant further distinguishes level accuracy from transferable dynamics.

9.1 Scope and limitations

The contribution concerns an interpretable forecasting construction and a conditional mathematical result. Regression coefficients are not calibrated physical kernel weights or Hurst estimates. Strong feature collinearity, wide conditional bootstrap ranges and occasional zero-coefficient experts limit structural interpretation. Daily GK variance is a range proxy, whose measurement properties differ from those of latent or intraday variance.

All market comparisons are retrospective. Fitting through 2023 establishes causal information sets but does not preregister later architectural or inferential decisions. Holm families are retrospective, block intervals are nominal and unadjusted, and resampling omits full model-selection uncertainty. Reconstructed fitting details and origin-year assignment also differ from the saved implementation; historical score changes cannot be attributed solely to purging. Original GARCH-family fitting code was unavailable and those forecasts were not regenerated.

The theorem assumes solution existence, uniform composed-integrand regularity and the driver-specific bound (17). Manufactured tests use zero stochastic convolution; sharper endpoint analysis is needed for their observed weak-example order. The SPY exercise uses an ETF proxy and a short sample, and the saved VaR mapping uses a smoothed twenty-day scale. Prospective confirmation, direct one-day risk modelling and stochastic numerical verification are distinct further studies.

9.2 Conclusion

The paper contributes a reproducible method for learning and comparing fractional memory scales while retaining transparent mathematical structure. Sparse learning and probabilistic adaptation reveal when horizon-specific composition is useful, and the conditional error theorem states the numerical guarantee precisely. The audited comparisons and simple-baseline checks define the evidence for the construction's benefits and limits. The framework supplies a reproducible basis for evaluating fractional memory under a prospective protocol with modelling choices fixed in advance.

Data and Code Availability

The research archive contains the six NIFTY input files, complete reconstructed fitting code, input hashes, software versions, selected settings, 1,317 per-origin prediction records, ten validation-boundary audits, 54 NIFTY metric records, all 48 paired NIFTY comparisons, stability and sensitivity outputs, and four extended SPY comparisons. Saved historical forecasts, econometric comparisons and risk diagnostics are retained separately from the reconstruction. Data redistribution remains subject to provider terms and journal policy.

References

- [1] F. Black and M. Scholes, "The pricing of options and corporate liabilities," *Journal of Political Economy*, vol. 81, no. 3, pp. 637–654, 1973.
- [2] R. C. Merton, "Theory of rational option pricing," *Bell Journal of Economics and Management Science*, vol. 4, no. 1, pp. 141–183, 1973.
- [3] F. Comte and E. Renault, "Long memory in continuous-time stochastic volatility models," *Mathematical Finance*, vol. 8, no. 4, pp. 291–323, 1998.
- [4] J. Gatheral, T. Jaisson, and M. Rosenbaum, "Volatility is rough," *Quantitative Finance*, vol. 18, no. 6, pp. 933–949, 2018.
- [5] M. Caputo, "Linear models of dissipation whose Q is almost frequency independent-II," *Geophysical Journal International*, vol. 13, no. 5, pp. 529–539, 1967.
- [6] I. Podlubny, *Fractional Differential Equations*. Academic Press, 1999.
- [7] B. B. Mandelbrot and J. W. Van Ness, "Fractional Brownian motions, fractional noises and applications," *SIAM Review*, vol. 10, no. 4, pp. 422–437, 1968.
- [8] E. Abi Jaber, M. Larsson, and S. Pulido, "Affine Volterra processes," *Annals of Applied Probability*, vol. 29, no. 5, pp. 3155–3200, 2019.
- [9] T. Bollerslev, "Generalized autoregressive conditional heteroskedasticity," *Journal of Econometrics*, vol. 31, no. 3, pp. 307–327, 1986.
- [10] D. B. Nelson, "Conditional heteroskedasticity in asset returns: A new approach," *Econometrica*, vol. 59, no. 2, pp. 347–370, 1991.
- [11] L. R. Glosten, R. Jagannathan, and D. E. Runkle, "On the relation between the expected value and the volatility of the nominal excess return on stocks," *Journal of Finance*, vol. 48, no. 5, pp. 1779–1801, 1993.
- [12] R. T. Baillie, T. Bollerslev, and H. O. Mikkelsen, "Fractionally integrated generalized autoregressive conditional heteroskedasticity," *Journal of Econometrics*, vol. 74, no. 1, pp. 3–30, 1996.
- [13] F. Corsi, "A simple approximate long-memory model of realized volatility," *Journal of Financial Econometrics*, vol. 7, no. 2, pp. 174–196, 2009.
- [14] F. X. Diebold and R. S. Mariano, "Comparing predictive accuracy," *Journal of Business & Economic Statistics*, vol. 13, no. 3, pp. 253–263, 1995.
- [15] P. H. Kupiec, "Techniques for verifying the accuracy of risk measurement models," *Journal of Derivatives*, vol. 3, no. 2, pp. 73–84, 1995.
- [16] P. F. Christoffersen, "Evaluating interval forecasts," *International Economic Review*, vol. 39, no. 4, pp. 841–862, 1998.
- [17] M. B. Garman and M. J. Klass, "On the estimation of security price volatilities from historical data," *The Journal of Business*, vol. 53, no. 1, pp. 67–78, 1980. doi: 10.1086/296072.
- [18] H. Zou and T. Hastie, "Regularization and variable selection via the elastic net," *Journal of the Royal Statistical Society: Series B*, vol. 67, no. 2, pp. 301–320, 2005. doi: 10.1111/j.1467-9868.2005.00503.x.

References continued

- [19] A. J. Patton, "Volatility forecast comparison using imperfect volatility proxies," *Journal of Econometrics*, vol. 160, no. 1, pp. 246–256, 2011. doi: 10.1016/j.jeconom.2010.03.034.
- [20] K. Christensen, M. V. Siggaard, and B. Veliyev, "A machine learning approach to volatility forecasting," *Journal of Financial Econometrics*, vol. 21, pp. 1680–1727, 2023. doi: 10.1093/jffinec/nbac020.
- [21] J. Guyon and J. Lekeufack, "Volatility is (mostly) path-dependent," *Quantitative Finance*, vol. 23, no. 9, pp. 1221–1258, 2023. doi: 10.1080/14697688.2023.2221281.
- [22] M. Fukasawa and T. Ugai, "Limit distributions for the discretization error of stochastic Volterra equations with fractional kernel," *The Annals of Applied Probability*, vol. 33, pp. 5071–5110, 2023. doi: 10.1214/23-AAP1941.
- [23] P. Gassiat, "Weak error rates of numerical schemes for rough volatility," *SIAM Journal on Financial Mathematics*, vol. 14, pp. 475–496, 2023. doi: 10.1137/22M1485760.
- [24] R. Cont and P. Das, "Rough volatility: Fact or artefact?" *Sankhya B*, vol. 86, pp. 191–223, 2024. doi: 10.1007/s13571-024-00322-2.
- [25] F. Biagini, L. Gonon, and N. Walter, "Approximation rates for deep calibration of (rough) stochastic volatility models," *SIAM Journal on Financial Mathematics*, vol. 15, no. 3, pp. 734–784, 2024. doi: 10.1137/23M1606769.
- [26] J. Luo, T. Klein, T. Walther, and Q. Ji, "Forecasting realized volatility of crude oil futures prices based on machine learning," *Journal of Forecasting*, vol. 43, no. 5, pp. 1422–1446, 2024. doi: 10.1002/for.3077.
- [27] Z.-C. Li, C. Xie, G.-J. Wang, Y. Zhu, Z.-J. Zeng, and J. Gong, "Forecasting global stock market volatilities: A shrinkage heterogeneous autoregressive (HAR) model with a large cross-market predictor set," *International Review of Economics & Finance*, vol. 93, pp. 673–711, 2024. doi: 10.1016/j.iref.2024.05.008.
- [28] R. R. Branco, A. Rubesam, and M. Zevallos, "Forecasting realized volatility: Does anything beat linear models?" *Journal of Empirical Finance*, vol. 78, Art. 101524, 2024. doi: 10.1016/j.jempfin.2024.101524.
- [29] J. D. Díaz, E. Hansen, and G. Cabrera, "Machine-learning stock market volatility: Predictability, drivers, and economic value," *International Review of Financial Analysis*, vol. 94, Art. 103286, 2024. doi: 10.1016/j.irfa.2024.103286.
- [30] E. Abi Jaber, C. Illand, and S. (Xiaoyuan) Li, "Joint SPX & VIX calibration with Gaussian polynomial volatility models: Deep pricing with quantization hints," *Mathematical Finance*, vol. 35, no. 2, pp. 470–519, 2025. doi: 10.1111/mafi.12451.
- [31] P. Bras and M. Fukasawa, "Weak error rates for numerical schemes of nonsingular stochastic Volterra equations with application to stochastic volatility models," *SIAM Journal on Financial Mathematics*, vol. 16, no. 1, pp. 1–28, 2025. doi: 10.1137/23M155671X.
- [32] Y. Ding, D. Kambouroudis, and D. G. McMillan, "Forecasting realised volatility using regime-switching models," *International Review of Economics & Finance*, vol. 101, Art. 104171, 2025. doi: 10.1016/j.iref.2025.104171.
- [33] R. Kiliç, "Linear and nonlinear econometric models versus machine-learning models: Evidence from realized-volatility forecasting," *Finance and Economics Discussion Series*, Board of Governors of the Federal Reserve System, no. 2025-061, revised Sept. 2026. doi: 10.17016/FEDS.2025.061r1.
- [34] S. Gao, L. Lu, J. Yan, Y. Tang, and J. Huang, "Numerical analysis of stochastic Volterra equations with singular kernels via the Euler–Maruyama scheme and large deviations," *Communications in Nonlinear Science and Numerical Simulation*, vol. 158, Art. 109827, 2026. doi: 10.1016/j.cnsns.2026.109827.
- [35] S. Holm, "A simple sequentially rejective multiple test procedure," *Scandinavian Journal of Statistics*, vol. 6, no. 2, pp. 65–70, 1979.