SvedbergOpen
DISSEMINATION OF KNOWLEDGE

International Journal of Artificial Intelligence and Machine Learning

Publisher's Home Page: https://www.svedbergopen.com/

Research Paper Open Access

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

Gauri N. Phad[1]*, V. R. Nikam[1]

[1]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
Volatility persistence is inherently multi-scale, yet most fractional financial models impose the form and strength of memory before observing the data. This paper develops an AI-assisted fractional stochastic Volterra framework in which an interpretable dictionary of power-law memory kernels is learned sparsely and, when supported by the data, adapted across latent volatility regimes. Daily NIFTY 50 open–high–low–close observations from 2021–2025 are used, with Garman–Klass variance as the primary range-based volatility proxy. Candidate memory exponents $\beta \in \{0.1, \ldots, 0.9\}$ generate causal fractional-memory features, while a positive Elastic Net selects the active scales. A three-state Gaussian mixture model provides probabilistic calm, normal and stress regimes, and regime-specific fractional experts are combined by soft posterior gating. The framework is supplemented by a paper-specific product-integration consistency and convergence analysis, together with manufactured-solution verification. In the 2024 NIFTY 50 internal holdout, regime adaptation is detrimental at the 5-day horizon but materially improves the 20-day specification, reducing QLIKE from approximately 0.3650 for the static AI kernel to 0.2795. Against GARCH, GJR-GARCH, EGARCH, FIGARCH, HAR, EWMA and historical-volatility benchmarks, the 20-day fractional model attains the second-lowest volatility RMSE, although HAR remains superior under QLIKE. A supplementary zero-shot NIFTY-to-SPY transfer in 2026 yields the lowest RMSE and MAE among the tested SPY models, and the resulting risk scale produces well-calibrated 95% and 99% VaR backtests. The evidence therefore supports a horizon-dependent and regime-dependent interpretation of financial memory rather than universal forecasting dominance.

Keywords: fractional stochastic Volterra equation; multi-scale memory; artificial intelligence; volatility forecasting; Garman–Klass variance; regime adaptation; Elastic Net; Value-at-Risk.

Introduction

Volatility forecasting is a central problem in mathematical finance because market uncertainty is neither constant nor memoryless. Financial returns display volatility clustering, persistent dependence and changing responses across short, medium and long horizons. Classical diffusion models such as geometric Brownian motion and the Black–Scholes–Merton framework are ana-lytically tractable, but their Markovian structure does not explicitly represent hereditary effects extending over multiple time scales [1, 2]. This limitation has motivated long-memory stochastic volatility, fractional Brownian motion, rough-volatility and stochastic Volterra approaches, all of which allow past information to influence present dynamics through non-local dependence [3, 7, 4, 8].

Fractional models are particularly attractive because power-law kernels provide a parsimonious mathematical representation of persistence. Their main modelling difficulty, however, is also their main source of flexibility: the memory order and kernel structure must be specified. Many studies therefore begin by fixing a fractional order, a Hurst parameter or a single power-law kernel and subsequently estimating only a small set of coefficients. Such a specification can be too restrictive if the effective memory scale changes with forecast horizon. At the opposite extreme, adding several fractional components without regularization can increase complexity without delivering better out-of-sample forecasts.

A second difficulty is that financial memory may be state dependent. Calm markets, ordinary trading conditions and stress episodes need not assign the same importance to recent and distant volatility observations. A fixed kernel forces these different regimes to share one memory structure. Regime-switching models address state dependence, while machine-learning methods can discover nonlinear or time-varying patterns, but these approaches are often separated from the fractional kernel itself. The resulting models can either preserve mathematical interpretability but remain inflexible, or gain predictive flexibility at the cost of a less transparent memory mechanism.

This paper develops a middle path: the memory mechanism remains explicitly fractional and interpretable, while sparse machine learning is used to determine which memory scales are active. We start from the finite kernel dictionary

$$K_{\Theta}(u) = \sum_{m=1}^{M} \lambda_m \frac{u^{\theta_m - 1}}{\Gamma(\theta_m)}, \qquad 0 < \theta_1 < \cdots < \theta_M < 1, \qquad \lambda_m \geq 0,$$

and estimate the non-negative weights using a positive Elastic Net. The dictionary therefore has a direct mathematical interpretation, while sparsity discourages unnecessary memory components. A probabilistic Gaussian-mixture regime layer is then used to determine whether separate fractional-memory experts are useful in calm, normal and stress states.

The empirical design is deliberately chronological. NIFTY 50 daily OHLC data from 2021–2025 are transformed into range-based volatility measures, with Garman–Klass variance used as the primary target. Model selection is restricted to information available through 2023, calendar year 2024 is used as the internal redesign holdout, and the previously examined 2025 period is not reused as a pristine confirmatory sample. This separation is important because the purpose of the redesigned study is not to reproduce an earlier in-sample improvement, but to evaluate whether adaptive fractional memory survives strong out-of-sample comparisons.

The study is organized around three research questions. First, do the useful fractional memory scales change with forecast horizon? Second, does probabilistic regime adaptation improve a sparse multi-scale fractional kernel, and if so, at which horizon? Third, does the resulting architecture remain useful when compared with strong econometric benchmarks and when transferred to an external market and a risk-management task? These questions shift the emphasis from proving that a more complex fractional model is always better to identifying the conditions under which additional memory complexity is justified.

The main contributions are as follows:

Data-driven fractional kernel learning. An interpretable finite dictionary of power-law kernels is combined with a positive Elastic Net, allowing the data to select active fractional scales while preserving non-negative memory weights.

Regime-adaptive memory. Separate sparse fractional experts are combined through Gaussian-mixture posterior probabilities, producing a soft state-dependent memory structure rather than a hard regime switch.

Numerical analysis. A product-integration discretization is studied for the learned finite-kernel Volterra equation. The resulting finite-horizon error bound separates deterministic regularity from stochastic-convolution approximation error, and manufactured-solution tests verify the expected convergence behaviour.

Strong leakage-controlled benchmarking. The proposed models are evaluated against GARCH, GJR-GARCH, EGARCH, FIGARCH, HAR, EWMA and historical volatility under a chronological design using QLIKE, RMSE, MAE and forecast-comparison tests.

Transfer and risk evaluation. A supplementary zero-shot NIFTY-to-SPY experiment examines cross-market robustness, and the transferred volatility signal is further assessed through 95% and 99% VaR backtesting with Kupiec and Christoffersen diagnostics.

Figure 1 summarizes the complete modelling and validation pipeline, from range-based volatility construction through sparse fractional-memory learning, regime adaptation, external transfer and risk evaluation.

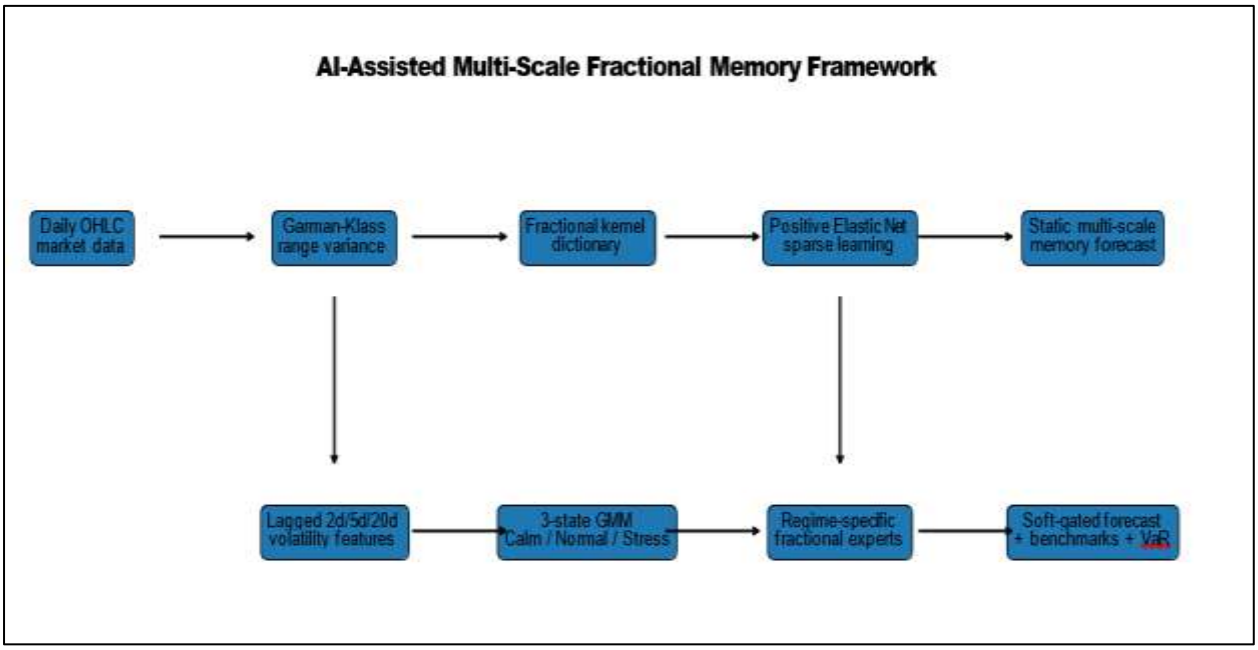

Figure 1: Overall AI-assisted multi-scale fractional-memory workflow. The empirical layer preserves an interpretable power-law memory basis while sparse learning and probabilistic regime gating determine which scales are useful for forecasting.

The central result is not that more fractional memory is always better. At the 5-day horizon, the data favour an almost single-scale short-memory specification and regime adaptation worsens performance. At the 20-day horizon, a broader memory mixture becomes useful and soft regime adaptation improves the static fractional specification. Even there, the model does not dominate every benchmark: HAR remains superior under QLIKE, while the fractional model performs strongly in volatility RMSE, cross-market transfer and risk calibration. This pattern motivates the paper's principal conclusion that financial memory should be treated as horizon dependent and potentially regime dependent.

The remainder of the paper is organized as follows. Section 2 reviews related work in fractional stochastic modelling and volatility forecasting. Sections 3 and 4 introduce the multi-scale fractional Volterra framework and the AI-based regime mechanism. Section 5 presents the numerical scheme and convergence result. Section 6 describes the data and leakage controls. Sections 7–9 report kernel-learning, regime and benchmark results. Sections 10 and 11 present the supplementary cross-market and risk-management applications. Sections 12–14 discuss the findings, limitations and conclusions.

Related Work

The proposed framework lies at the intersection of fractional stochastic modelling, path-dependent volatility, volatility econometrics and data-driven model selection. Recent work shows that these literatures increasingly overlap, but they still treat the composition of memory in different ways.

Fractional, rough and Volterra representations of memory

Fractional calculus introduces non-local operators whose present value depends on a history of past states. For $\gamma > 0$, the Riemann–Liouville fractional integral is

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

The power-law kernel in (1) provides a natural representation of hereditary dependence [5, 6]. Fractional Brownian motion provides a distinct stochastic source of dependence through correlated increments controlled by the Hurst parameter [7]. Long-memory stochastic volatility [3], rough volatility [4], and stochastic Volterra models [8] exploit these ideas in different ways.

Recent research also cautions against treating empirical persistence diagnostics as definitive evidence about latent spot volatility. Cont and Das [24] show that apparent roughness in realized-volatility signals can be affected by the observation and estimation procedure. For this reason, the autocorrelation and related dependence diagnostics in the present paper are interpreted descriptively for daily range-based variance, not as proof that latent spot volatility follows a unique rough fractional process.

The Volterra literature has meanwhile moved toward flexible kernel comparison and calibra-tion. Abi Jaber [30] compares Markovian and non-Markovian kernels in a Gaussian Volterra volatility class, while Biagini, Gonon and Walter [25] study deep approximation for calibration of classical and rough stochastic-volatility models. These contributions use data or machine learning to calibrate structured stochastic models, but they do not address the same forecasting question considered here: which members of an interpretable fractional power-law dictionary are empirically useful at a given horizon and regime.

A closely related empirical perspective is path-dependent volatility. Guyon and Lekeufack [21] show that simple weighted functions of historical returns and squared returns can explain a substantial part of future and implied volatility, with power-law-like path dependence playing a central role. Their results strengthen the motivation for learning memory weights from data rather than fixing one universal decay rate.

Machine learning and econometric volatility forecasting

The econometric benchmark literature models persistence through recursive conditional-variance dynamics. GARCH captures persistent conditional heteroskedasticity [9]; EGARCH and GJR-GARCH incorporate asymmetric responses [10, 11]; FIGARCH introduces fractional integration into conditional variance [12]; and HAR represents heterogeneous dependence across daily, weekly and monthly scales [13].

Machine learning has increasingly been used to extend or compete with these models. Christensen, Siggaard and Veliyev [20] show that regularization, tree methods and neural networks can be competitive for realized-volatility forecasting, with gains that vary by horizon. Luo et al. [26] similarly report horizon-dependent variable-selection patterns in machine-learning forecasts of crude-oil realized volatility. Li et al. [27] use LASSO and Elastic Net in a high-dimensional HAR setting and find that the useful predictor set changes with forecast horizon. At the same time, broad international evidence indicates that nonlinear or high-capacity methods do not systematically dominate simpler linear benchmarks across all horizons and evaluation criteria [28, 29]. A very recent Federal Reserve study reaches a closely related conclusion: the best representation of persistence and nonlinearity depends materially on forecast horizon, with regime-switching, long-memory and machine-learning models ranking differently across horizons [33]. These findings support the conservative evaluation principle adopted in the present paper: additional model complexity must earn its place out of sample.

Regime dependence provides another route to nonlinearity. Ding, Kambouroudis and McMillan [32] show that regime-switching extensions of AR/HAR volatility models can improve forecasting and risk-management performance, although the benefit is horizon dependent. The present study differs by allowing regimes to alter the fractional memory composition itself, rather than only switching coefficients in a conventional autoregressive volatility equation.

Gap and position of the present study

The motivating gap is therefore narrower than the statement that "financial volatility has memory." That fact is already well established. The unresolved modelling choice is how to retain a mathematically interpretable power-law memory structure while allowing the relevant scales to be learned, sparsified and, only when justified, adapted across regimes.

The machine-learning component in this paper does not replace the stochastic Volterra model with a black-box predictor. Instead, it estimates the empirical contribution of a finite fractional-memory basis. The stochastic Volterra equation supplies the structural non-local model and the numerical-analysis object; the forecasting layer supplies a causal reduced-form projection of the same power-law family onto observable daily volatility. Table 1 summarizes the positioning relative to representative recent studies. The table is not intended as an exhaustive literature survey; it isolates the dimensions most relevant to the contribution claimed here.

Table 1: Representative literature positioning relative to the present study.

Study	**Memory Representation**	**Data-Driven Memory Learning**	**Regime Adaptation**
Gatheral et al. [4]	Rough fractional volatility	Fixed structural roughness/kernel	No
Guyon–Lekeufack [21]	Path-dependent power-law weighting of past market information	Empirically fitted parametric weights	No
Christensen et al. [20]	HAR/ML representation of volatility persistence	Predictor learning and regularization	No
Abi Jaber [30]	Gaussian Volterra volatility with alternative kernels	Kernel calibration and comparison	No
Ding et al. [32]	AR/HAR volatility with state switching	No fractional-memory dictionary	Yes
Present study	Finite fractional Volterra power-law dictionary	Sparse non-negative learning of active memory scales	**Yes; soft probabilistic gating**

Fractional Stochastic Volterra Framework

Continuous-time multi-scale memory model

Let $(\Omega, F, \{F_t\}_{t\geq 0}, P)$ be a filtered probability space and let $X(t)$ denote the state variable. We consider the finite multi-scale fractional Volterra representation

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

where $0 < \alpha < 1$, $0 < \beta_m < 1$, $\lambda_m \geq 0$, and $\Xi(t)$ denotes an additive stochastic convolution.

The corresponding finite memory kernel is

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

Thus, for an observable process V,

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

Equation (4) gives the key interpretation used in the empirical layer. Each candidate exponent β_m defines one power-law memory channel. Small β_m gives a more rapidly decaying weight sequence and therefore concentrates relatively more mass on recent observations. Larger β_m produces slower decay and represents longer effective memory.

The parameter α and the exponents β_m play different roles. The order α belongs to the continuous-time fractional Volterra backbone, while the β_m values index the candidate hereditary scales whose empirical importance is learned from observed volatility. Keeping these roles distinct avoids interpreting the data-driven kernel dictionary as a direct estimator of every parameter in the stochastic integral equation.

From the Volterra kernel to causal volatility-memory features

Let V_t denote daily Garman–Klass variance. The empirical forecasting layer uses the continuous-time power-law structure as a reduced-form, causal approximation to the fractional convolution. For a lookback window L, define

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

The normalization makes each Mβ(t) comparable across candidate exponents and isolates the shape of the memory decay from its overall regression coefficient. We use L = 252, corresponding approximately to one trading year.
The weights in (5) preserve the regularly varying power-law form of the fractional kernel. They are not the same object as the product-integration weights used later for numerical solution of the Volterra equation. This distinction is deliberate: the numerical scheme approximates the continuous-time integral equation, whereas the forecasting layer constructs interpretable discrete memory features from observed daily volatility.
For horizon h, the empirical target is

$$Y_{t,h} = \log\left(\frac{1}{h}\sum_{j=1}^{h} V_{t+j} + \varepsilon\right) \tag{6}$$

with ε = 10−12 used only to avoid evaluating the logarithm at zero.

Sparse learning of fractional memory scales
Let

⊤ Zt = [log(Mβ1 (t) + ε), . . . , log(MβM (t) + ε)]

denote the standardized fractional-memory feature vector. The static AI kernel is estimated by a positive Elastic Net, following the regularization principle of Zou and Hastie [18]:

$$(\hat b_0, \hat b) = \arg\min_{b_0, b\geq 0} \left\{ \frac{1}{N}\sum_{t=1}^{N} \left(Y_{t,h} - b_0 - Z_t^\top b\right)^2 + \tau\left(\eta_E\|b\|_1 + \frac{1-\eta_E}{2}\|b\|_2^2\right)\right\},$$

where τ > 0 controls the total regularization strength and ηE ∈ [0, 1] controls the ℓ1/ℓ2 mixture. The notation τ and ηE is used here to avoid confusion with the fractional-kernel coefficients λm and the fractional order α. Non-negativity is imposed for interpretability: an active coefficient indicates that the corre-sponding memory scale contributes positively to the forecasted volatility level. After transforming coefficients back to the original feature scale, normalized memory importance is defined by

$$\omega_m = \frac{\hat b_m}{\sum_{\ell:\hat b_\ell>0} \hat b_\ell}, \quad \sum_m \omega_m = 1, \tag{8}$$

for active coefficients. An effective memory exponent can then be summarized as
This scalar is used only as a summary; the full vector (ω1, . . . , ωM) retains the multi-scale structure.
Because the empirical model is fitted on the logarithmic target, forecasts are transformed back to the variance scale using the standard lognormal residual correction

$$\beta_{\mathrm{eff}} = \sum_m \omega_m \beta_m. \tag{9}$$

$$\hat V_{t,h} = \exp\left(\hat Y_{t,h} + \frac{1}{2}\hat s_h^2\right) - \varepsilon, \tag{10}$$

where s2 h is the training residual variance.

Interpretation of the mathematical–empirical connection
The continuous-time equation (2) and the empirical regression layer serve complementary purposes. Equation (2) supplies the structured stochastic-memory model and the numerical-analysis object. Equations (5)–(10) provide an observable reduced-form method for learning which fractional kernel shapes are useful for volatility forecasting.
Accordingly, the Elastic-Net coefficients should not be interpreted as direct pathwise estimates of all coefficients in (2). Their role is narrower and more defensible: they identify the empirical importance of a finite fractional-memory basis derived from the same power-law kernel family. This separation allows the paper to combine mathematical structure and AI-based learning without conflating stochastic-equation estimation with feature-based forecasting.

AI-Based Regime Adaptation

Regime identification

The static sparse kernel assumes that one learned memory composition applies throughout the sample. To test whether the relevant memory scales change with market state, we introduce a three-component Gaussian mixture model using only information available at the forecast origin.

Define

$$R_t = \left(\log \bar{V}_{t,2}, \log \bar{V}_{t,5}, \log \bar{V}_{t,20}\right)^\top, \qquad (11)$$

where $\bar{V}_{t,k}$ is the trailing k-day average Garman–Klass variance. The feature vector is standardized using training-period moments, and the Gaussian mixture is fitted on the training period only. Components are ordered by their training-period 20-day volatility level and labelled Calm, Normal and Stress.

Let

$$\pi_r(t) = P(R_t \text{ belongs to regime } r \mid F_t), \qquad r = 1, 2, 3,$$

denote the posterior regime probabilities. These probabilities are observable at the forecast origin because every element of R_t is lagged.

Regime-specific fractional experts

A separate positive Elastic-Net fractional-memory expert is fitted within each training-period regime using the same candidate β dictionary and the same horizon-specific regularization setting selected for the static model. Holding the regularization structure fixed isolates the value of regime adaptation from additional hyperparameter search.

If $\hat{V}^{(r)}_{t,h}$ denotes the forecast from regime expert r, the hard-gated forecast is

$$\hat{V}^{hard}_{t,h} = \hat{V}^{(\hat{r}_t)}_{t,h}, \qquad \hat{r}_t = \arg\max_r \pi_r(t), \qquad (12)$$

whereas the primary soft-gated forecast is

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

Soft gating is preferred as the main adaptive specification because it propagates regime uncertainty instead of forcing every observation into a single state. The empirical question is then whether this additional state dependence improves genuinely out-of-sample forecasts. If it does not, the static sparse kernel remains the more parsimonious model.

Numerical Approximation and Convergence

Numerical approximation of stochastic Volterra equations with singular or fractional kernels is an active research area, with recent results covering discretization errors, weak convergence and singular-kernel schemes [22, 23, 31, 34]. The purpose of this section is therefore not to claim a universal new stochastic-Volterra convergence theory. Instead, we give a paper-specific product-integration result for the finite learned kernel in (3) and make the stochastic-convolution approximation error explicit.

Let $t_n = nh$, $n = 0, \ldots, N$, and define

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

The product-integration approximation of (2) is

$$X_n = X_0 + h^\alpha \sum_{j=0}^{n-1} a_{n-j}^{(\alpha)} F(t_j, X_j) + \sum_{m=1}^{M} \lambda_m h^{\alpha+\beta_m} \sum_{j=0}^{n-1} a_{n-j}^{(\alpha+\beta_m)} G_m(X_j) + \Xi_n. \qquad (15)$$

Theorem 1 (Conditional finite-horizon grid error). Assume that a unique square-integrable solution of (2) exists on [0, T]. Suppose that F and Gm are globally Lipschitz in the state variable with constants LF and Lm, and that the composed functions

$$t \mapsto F(t, X(t)), \quad t \mapsto G_m(X(t))$$

are uniformly η-Hölder continuous on $[0, T]$ *in* L^2*, for some* $\eta \in (0, 1]$*. Define*

$$q_T = \frac{L_F T^{\alpha}}{\Gamma(\alpha+1)} + \sum_{m=1} \frac{|\lambda_m| L_m T^{\alpha+\beta_m}}{\Gamma(\alpha+\beta_m+1)}. \qquad (16)$$

If $q_T < 1$ *and the chosen approximation of the additive stochastic convolution satisfies*

$$\max_{0 \le n \le N} \|\Xi(t_n) - \Xi_n\|_{L^2} \le C_{\Xi} h^{\rho} \qquad (17)$$

for some $\rho > 0$*, then*

$$\max_{0 \le n \le N} \|X(t_n) - X_n\|_{L^2} \le C\,(h^{\eta} + h^{\rho}), \qquad (18)$$

where C is independent of h.

Proof. For an η-Hölder integrand ϕ, product integration with left-endpoint interpolation gives

$$\left\| I^{\nu}\phi(t_n) - h^{\nu} \sum_{j=0}^{n-1} a^{(\nu)}_{n-j}\phi(t_j) \right\|_{L_2} \le C_{\phi} h^{\eta} \frac{t^{\nu}}{\Gamma(\nu+1)}.$$

Subtracting (15) from the exact integral equation, using the Lipschitz assumptions and taking the maximum grid error

$$E_N = \max_{0 \le n \le N} \|X(t_n) - X_n\|_{L^2},$$

yields

$$E_N \le C_0 h^{\eta} + C_{\Xi} h^{\rho} + q_T E_N.$$

Since $q_T < 1$, division by $1 - q_T$ gives (18).

The theorem is intentionally conditional on (17). The value of ρ depends on the stochastic driver, integration convention and quadrature. This is important in view of the distinct rates established for different stochastic Volterra settings in the recent numerical literature [22, 23, 31, 34]. We therefore do not insert a universal fractional-Brownian-motion rate into (18).

For manufactured-solution verification, we use the deterministic specialization

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

If $X(t) = t^{\nu}$, then

A representative numerical-analysis configuration uses α = 0.7, β = (0.1, 0.2, 0.3) and scaled non-negative coefficients that preserve the relative 20-day learned memory proportions while keeping the illustrative finite-horizon contraction factor below one. These coefficients are used only for numerical verification; they are not presented as new empirical parameter estimates. The resulting q_T is approximately 0.4147 for T = 1. Across meshes N = 25, . . . , 1600, the observed maximum-norm order is approximately 1.002 for the smooth solution $X(t) = t^2$ and 0.800 for $X(t) = t^{0.8}$, illustrating the expected regularity-limited behaviour.

Figure 2 visualizes the corresponding log–log error decay. The separation between the smooth and weakly regular manufactured solutions makes the regularity-limited character of the deterministic discretization error explicit.

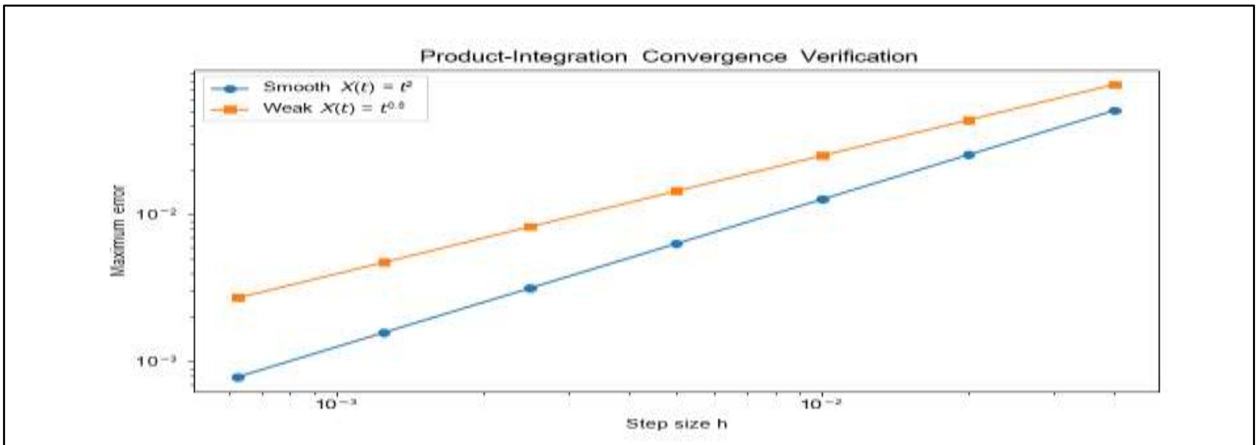

Figure 2: Product-integration convergence verification for smooth and weakly regular manufac-tured solutions.

Data and Experimental Design

NIFTY 50 data and volatility proxy

The main empirical data set contains 1,239 daily NIFTY 50 open, high, low and close (OHLC) observations from 1 January 2021 to 31 December 2025. The primary volatility proxy is the Garman–Klass range-based variance estimator [17],

$$V_t^{GK} = \frac{1}{2} \quad \log \quad \frac{H_t}{L_t} \quad ^2 \quad - (2 \log 2 - 1) \quad \log \quad \frac{C_t}{O_t} \quad ^2 \quad . \ (21)$$

Parkinson and Rogers–Satchell range estimators were also examined during target construction, but Garman–Klass displayed the strongest training-period persistence diagnostics and is retained as the primary target. The term "realized variance" is not used for (21): the data are daily OHLC observations rather than intraday high-frequency returns. The object forecast in this paper is therefore a range-based daily variance proxy.

Chronological development protocol

The redesigned study uses a strict chronological protocol summarized in Table 2. The model class, fractional dictionary and hyperparameter-selection rules are developed using target dates no later than 31 December 2023. Calendar year 2024 is the internal holdout used to compare static, regime-adaptive and econometric benchmark specifications. The NIFTY 50 observations from 2025 had been inspected in earlier model-development work and are therefore not labelled as a pristine confirmatory test in the redesigned paper.
For the sparse-kernel stage, the 5-day and 20-day model-selection samples contain 486 and 471 targets, respectively, through 2023; calendar 2024 contains 248 holdout targets for each horizon. Five-fold expanding time-series cross-validation is used rather than random folds.

Forecast losses and leakage controls

QLIKE is the primary volatility loss because it belongs to the class of losses robust to noise in volatility proxies discussed by Patton [19]. Secondary evaluation uses annualized-volatility

Table 2: Chronological role of the data in the redesigned study.

Data Block	Role	Main Restriction
NIFTY 50 through 2023	Model development and expanding-window cross-validation	Kernel dictionary, regularization, and regime architecture selected without using 2024 information
NIFTY 50, 2024	Internal out-of-sample holdout	Used as an internal out-of-sample evaluation period
NIFTY 50, 2025	Historical/development evidence only	Used as historical/development evidence and not treated as a pristine final test set
SPY, 2026	Supplementary zero-shot robustness experiment	Used for architecture and benchmark comparison; not reused for further tuning after the reported redesign. Not considered a pristine final test because it had been examined in earlier work. Fractional expert coefficients were transferred from NIFTY without SPY-specific refitting.

Used for architecture and benchmark compar-ison; not reused for further tuning after the reported redesign
Not called a pristine final test because it had been examined in earlier work
Fractional expert coefficients transferred from NIFTY without SPY-specific refitting

RMSE, MAE and forecast correlation. Pairwise predictive-accuracy comparisons use Diebold–Mariano/HAC statistics where appropriate [14].

The following controls are imposed throughout:

- observations are never randomly shuffled;
- fractional memory features use only lags available at the forecast origin;
- regime features and posterior probabilities use only trailing volatility information;
- GMM standardization moments and regime labels are estimated from training information only;
- regularization settings are selected before the 2024 holdout is evaluated;
- the 2024 holdout is reported as internal model-development evidence, not as untouched final confirmation;
- the SPY experiment is explicitly labelled supplementary because SPY is an ETF proxy rather than the S&P 500 index itself.

Pre-2024 diagnostics show substantial persistence in the daily range-based variance. For Garman–Klass variance over 2021–2023, the lag-one autocorrelation is approximately 0.359, and the Ljung–Box test at 20 lags strongly rejects serial independence (p < 10−100). Figure 3 shows that positive dependence persists across multiple trading-day lags. Consistent with the caution raised by Cont and Das [24], these statistics are interpreted as dependence diagnostics for the observed proxy rather than identification of a unique latent roughness parameter.

Static Multi-Scale Kernel Learning

The candidate fractional dictionary is $\beta \in \{0.1, 0.2, \ldots, 0.9\}$.

Five-fold expanding-window time-series cross-validation is used, with QLIKE as the primary selection loss. The use of sparse regularization is consistent with recent volatility-forecasting evidence that predictor relevance and optimal shrinkage can vary materially across horizons [20, 26, 27].

At the 5-day horizon, the positive Elastic Net places approximately 98.3% of normalized coefficient mass on $\beta = 0.1$ and 1.7% on $\beta = 0.2$, yielding an effective memory exponent near 0.102. The 2024 QLIKE is 0.4784, only slightly better than the single-scale benchmark.

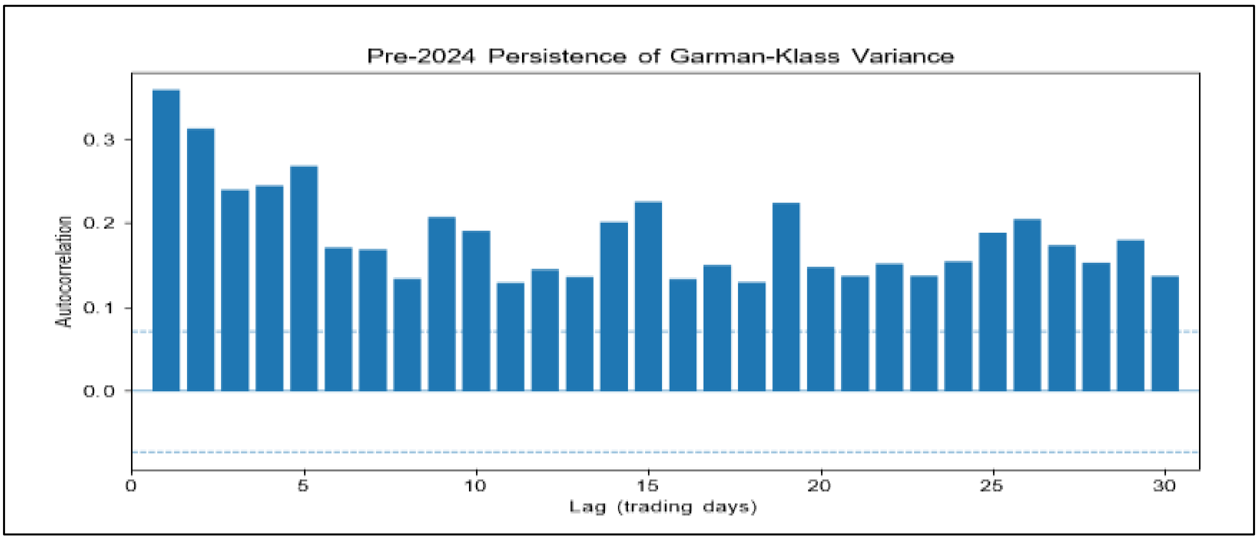

Figure 3: Autocorrelation of pre-2024 Garman–Klass variance for lags 1–30. Dashed lines show the approximate 95% white-noise bounds.

At the 20-day horizon, the selected multi-scale solution spreads its active coefficient mass across $\beta = 0.1$, 0.2 and 0.3, with approximate normalized shares 42.3%, 14.0% and 43.7%. The effective exponent is approximately 0.201. This supports the interpretation that the relevant fractional-memory scale broadens with forecast horizon.

The contrast is visible in Figure 4: the 5-day solution is nearly degenerate at the shortest candidate scale, whereas the 20-day solution distributes importance across three nearby scales.

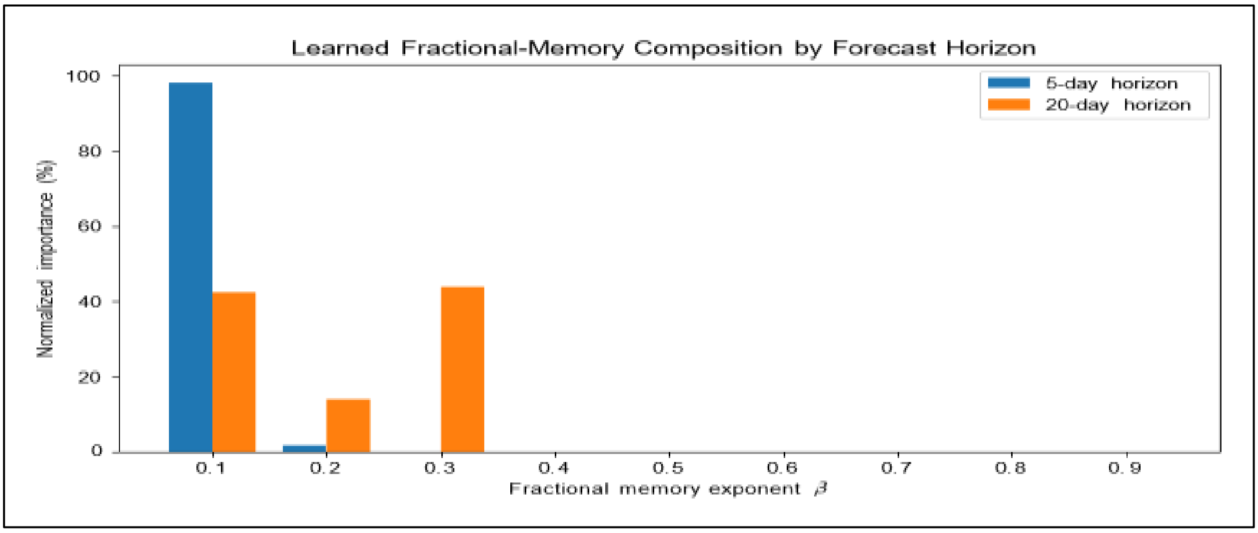

Figure 4: Normalized learned fractional-memory importance by forecast horizon.

Regime-Adaptive Results

Table 3 reports the internal 2024 comparison.

At 5 days, regime adaptation worsens QLIKE relative to the static sparse kernel, and the de-terioration is statistically significant. At 20 days, the soft regime-adaptive model reduces QLIKE from approximately 0.3650 to 0.2795, a relative reduction of about 23.4%. The corresponding

Table 3: 2024 internal-holdout QLIKE for static and regime-adaptive models.

Model	5-day QLIKE	20-day QLIKE
Fixed dual-memory	0.480285	0.364813
Static AI multi-scale	0.478432	0.364997
Hard regime-adaptive	0.514676	0.306877
Soft regime-adaptive	0.519351	0.279488

Diebold–Mariano comparison against the static AI model is not significant at the 5% level, so the result is treated as a medium-horizon hypothesis rather than proof of universal superiority. Figure 5 provides the key interpretability result for the 20-day model. The calm-state expert places its active weight on the long candidate scale β = 0.9, while the normal and stress experts concentrate on the short β = 0.1–0.2 region. Thus the regime mechanism changes the composition of fractional memory rather than merely adding an unrestricted nonlinear adjustment.

Regime-Specific Fractional-Memory Composition at 20 Days

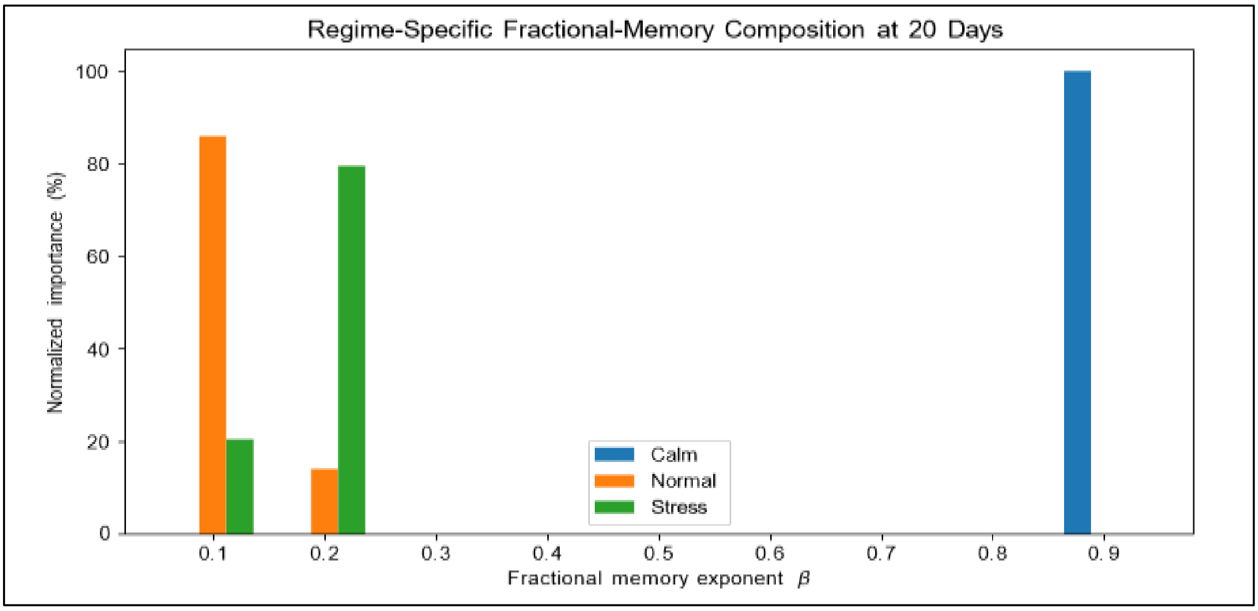

Figure 5: Normalized regime-specific fractional-memory importance for the 20-day horizon.

Strong Econometric Benchmark Comparison

The 20-day fractional architecture is compared with GARCH, GJR-GARCH, EGARCH, FI-GARCH, HAR-GK, EWMA-GK and Historical-20 GK. All benchmark parameters are estimated using pre-2024 information.

The proposed fractional model has the second-lowest volatility RMSE and improves RMSE relative to GARCH, GJR-GARCH, EGARCH, FIGARCH, EWMA and Historical-20. HAR remains stronger under QLIKE, and the Diebold–Mariano test favors HAR significantly. Ac-cordingly, the paper does not claim uniform forecasting dominance.

Figure 6 compares the time path of actual 20-day annualized Garman–Klass volatility with the selected fractional model and two strong parsimonious benchmarks. The figure reinforces the mixed metric result: the fractional forecast is comparatively smooth and close in level over much of the holdout, while HAR and EWMA respond differently during volatility changes.

Table 4: 2024 internal-holdout benchmark comparison at the 20-day horizon.

Model	QLIKE	Annualized-volatility RMSE
HAR-GK	0.22127	0.04064
EGARCH (1,1)	0.22735	0.04796
GJR-GARCH (1,1)	0.26195	0.05893
EWMA-GK	0.27441	0.04657
Soft regime-adaptive fractional	0.27949	0.04267
GARCH (1,1)	0.29538	0.06180
Historical-20 GK	0.29784	0.05064
FIGARCH(1,d,1)	0.31281	0.06227

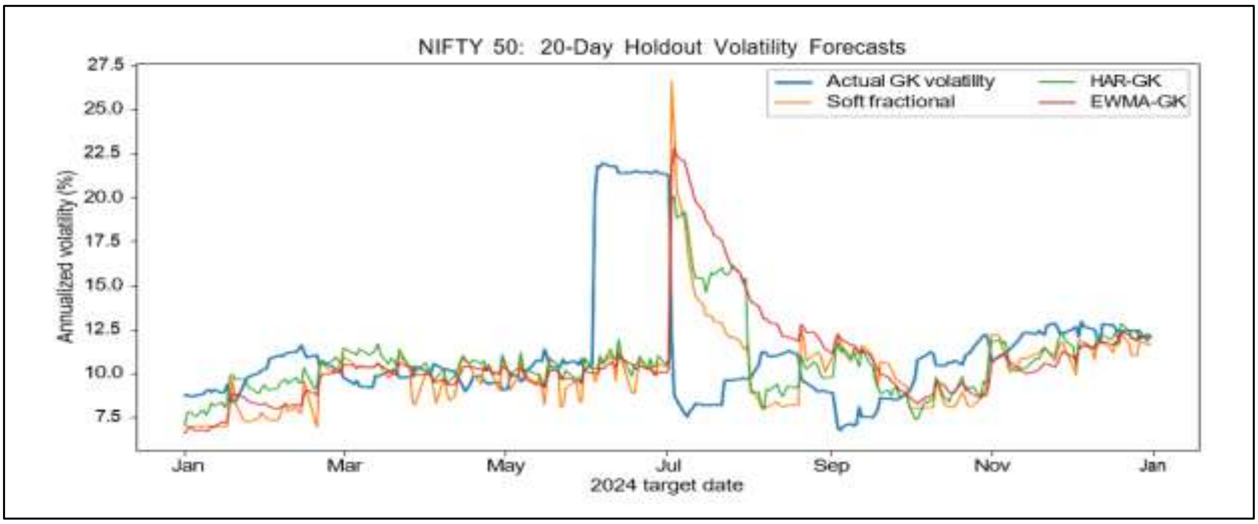

Figure 6: NIFTY 50 internal holdout: actual 20-day annualized Garman–Klass volatility and selected forecasts during 2024.

Supplementary Zero-Shot Cross-Market Validation

A supplementary external robustness experiment transfers the frozen 20-day fractional architec-ture from NIFTY 50 to SPY without re-estimating the fractional expert coefficients on SPY. The test contains 148 fully observed forecast origins in 2026.

Table 5: SPY 2026 zero-shot external validation.

Model	QLIKE	Ann. vol. RMSE	Ann. vol. MAE
Zero-shot fractional	0.19397	0.02952	0.02258
HAR-GK	0.16487	0.03655	0.03065
EWMA-GK	0.19034	0.03158	0.02743
Historical-20	0.26426	0.03519	0.02981

The zero-shot fractional model gives the lowest RMSE and MAE, while HAR provides the lowest QLIKE. Pairwise QLIKE differences between the transferred fractional model and HAR, EWMA and Historical-20 are not statistically significant. The result is interpreted as evidence of transfer robustness rather than dominance.
Figure 7 shows the corresponding forecast paths. The zero-shot fractional forecast remains stable across the 2026 sample; its low RMSE and MAE reflect level accuracy rather than strong short-run comovement, consistent with the weak forecast correlation reported in Table 5.

Figure 7: SPY 2026 zero-shot robustness experiment: actual 20-day annualized Garman–Klass volatility and selected forecasts.

Exploratory Financial Risk Application

To examine practical risk usefulness, the frozen 20-day expected average GK variance is used as a smoothed daily risk scale. Coverage is evaluated with the Kupiec and Christoffersen backtests [15, 16]. Because GK variance is range-based while VaR is evaluated against close-to-close returns, a fixed source-market scaling factor is estimated from NIFTY 50 observations from 2021–2023:

$$\kappa = \frac{\text{mean}(r_t^2)}{\text{mean}(V_t^{GK})} \approx 1.6502. \quad (22)$$

mean(r2)

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

For Gaussian innovations, the 95% and 99% loss VaR levels are constructed from the corresponding normal quantiles. A Student-t robustness specification uses $\nu \approx 5.39$ degrees of freedom estimated only from NIFTY 50 training returns.

Table 6: Gaussian VaR backtest on SPY 2026.

Model	95% exc.	Rate	Kupiec p	Indep. p	CC p	Mean VaR
Zero-shot fractional	8	5.41%	0.823	0.430	0.714	1.442%
HAR-GK	2	1.35%	0.016	0.814	0.054	1.747%
EWMA-GK	7	4.73%	0.879	0.316	0.598	1.471%
Historical-20	8	5.41%	0.823	0.337	0.615	1.456%

The fractional 95% VaR produces 8 exceptions out of 148 observations, corresponding to a 5.41% empirical exception rate. It passes the Kupiec unconditional-coverage test, the Christoffersen independence test and the conditional-coverage test. At 99%, the fractional model records 2 exceptions and again passes the coverage diagnostics. HAR is substantially more conservative at 95% and fails the Kupiec test. The fractional model also has the lowest 95% Gaussian quantile loss among the four tested risk specifications; the loss difference relative to HAR is statistically significant, while differences relative to EWMA and Historical-20 are not. Figure 8 plots the realized daily loss together with the fractional 95% and 99% VaR thresholds.

The relatively small number of threshold crossings is visually consistent with the formal coverage tests, while the figure also makes clear that the risk signal is smooth because it is inherited from the 20-day expected variance forecast.

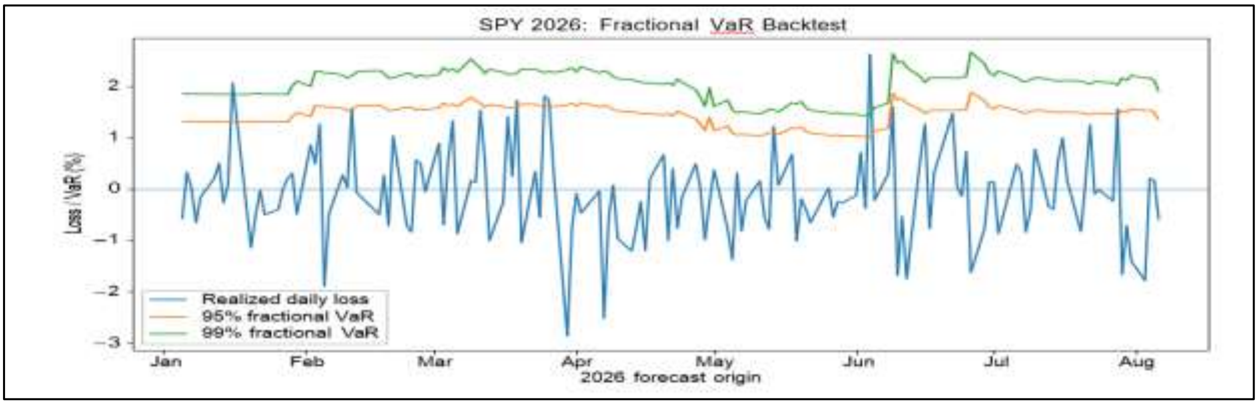

Figure 8: SPY 2026 exploratory VaR application using the frozen fractional volatility signal.

Discussion

The results support four substantive interpretations.

First, useful fractional memory is horizon dependent. At 5 days, almost all static-kernel importance is concentrated near $\beta = 0.1$, and the addition of regime-specific experts worsensthe holdout loss. At 20 days, the active fractional support broadens and soft regime adaptation materially improves the static specification. This is consistent with recent machine-learning and shrinkage studies reporting horizon-dependent predictor relevance and model rankings [20, 26, 27, 33]. The present contribution is more specific: it is the composition of an interpretable fractional-memory basis itself that changes with horizon and regime.

Second, state dependence is useful only conditionally. The 20-day regime experts reveal economically interpretable differences in memory composition, but the same mechanism fails at the shorter horizon. This result parallels the broader regime-switching literature, where gains are often sensitive to forecast horizon and evaluation criterion [32]. The negative 5-day result is therefore retained as part of the evidence rather than removed as an unsuccessful specification.

Third, no single forecasting model dominates every criterion. HAR is significantly better than the selected fractional model under 20-day QLIKE on the 2024 holdout, whereas the fractional model ranks second in annualized-volatility RMSE and performs strongly in the supplementary SPY transfer. This mixed ranking is consistent with recent international evidence that nonlinear and machine-learning models do not uniformly dominate simpler volatility benchmarks [28, 29]. The principal contribution is thus an interpretable adaptive-memory mechanism, not a universal winner claim.

Fourth, the cross-market and VaR results suggest that the learned memory representation has practical value beyond the original NIFTY holdout. Nevertheless, the zero-shot SPY exercise is supplementary and the VaR mapping uses a source-trained range-to-close variance bridge. These experiments strengthen the application case but should not be confused with a fully re-estimated one-day risk model or a four-index confirmatory study.

Limitations

Several limitations are material to the interpretation of the results.

The strongest fresh external experiment currently uses SPY as an S&P 500 market proxy rather than the underlying index. It is therefore supplementary robustness evidence, not the preregistered four-index confirmation.

Daily Garman–Klass variance is a range-based proxy constructed from OHLC data, not intraday realized variance. Measurement error in any volatility proxy can affect persistence diagnostics and forecast evaluation.

The SPY zero-shot forecast correlation is weak even though RMSE and MAE are favorable. Low average error therefore does not imply accurate tracking of every short-run volatility movement.

The stochastic-convolution rate in the numerical theorem is conditional on the chosen driver and quadrature. The paper does not claim a universal fBm convergence order.

The manufactured-solution coefficients are illustrative numerical-analysis parameters scaled to satisfy the finite-horizon contraction condition; they are not empirical estimates of the NIFTY forecasting equation.

The VaR application maps a 20-day expected average range-variance forecast into a smoothed one-day risk scale using a source-trained variance bridge. It should be interpreted as an exploratory risk application, not as a dedicated one-day VaR model.

NIFTY 50 data from 2025 were examined during earlier development and are not presented as pristine evidence for the redesigned architecture.

A stronger confirmatory package would apply the completely frozen architecture to actual 2026 OHLC data for NIFTY 50, S&P 500, FTSE 100 and Nikkei 225.

Conclusion

This paper develops an AI-assisted multi-scale fractional stochastic Volterra framework for financial volatility in which interpretable power-law memory components are learned rather than fixed in advance. Sparse non-negative learning identifies the relevant fractional scales, while probabilistic regime adaptation allows the memory composition to change with market conditions.

The evidence shows that additional fractional complexity is not uniformly beneficial. At the 5-day horizon, a simple short-memory representation is preferable. At the 20-day horizon, however, a soft regime-adaptive multi-scale kernel becomes competitive with established econometric volatility models and achieves strong RMSE performance. Product-integration analysis shows that the deterministic convergence rate is controlled by regularity, while the stochastic-convolution rate enters transparently through the overall mean-square error bound.

The supplementary NIFTY-to-SPY zero-shot experiment further indicates that the learned fractional-memory structure can transfer across markets without target-market refitting, and the resulting risk scale yields well-calibrated 95% and 99% VaR backtests. The overall contribution is therefore not a claim that fractional memory always dominates conventional models, but rather that adaptive, interpretable and horizon-specific fractional memory provides a useful mathematical and computational framework for volatility modelling and risk analysis.

Data and Code Availability

The NIFTY 50 empirical analysis uses daily OHLC index observations for 2021–2025. The supplementary SPY experiment uses daily OHLC/close data obtained through Financial Mod-eling Prep (FMP). Derived analysis scripts and workbooks were retained for reproducibility. Availability of raw and derived files is subject to the redistribution terms of the underlying data providers and the policy of the target journal.

References

- F. Black and M. Scholes, "The pricing of options and corporate liabilities," Journal of Political Economy, vol. 81, no. 3, pp. 637–654, 1973.
- R. C. Merton, "Theory of rational option pricing," Bell Journal of Economics and Manage-ment Science, vol. 4, no. 1, pp. 141–183, 1973.
- F. Comte and E. Renault, "Long memory in continuous-time stochastic volatility models," Mathematical Finance, vol. 8, no. 4, pp. 291–323, 1998.
- J. Gatheral, T. Jaisson, and M. Rosenbaum, "Volatility is rough," Quantitative Finance, vol. 18, no. 6, pp. 933–949, 2018.
- M. Caputo, "Linear models of dissipation whose Q is almost frequency independent–II," Geophysical Journal International, vol. 13, no. 5, pp. 529–539, 1967.

I. Podlubny, Fractional Differential Equations. Academic Press, 1999.

- 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.
- E. Abi Jaber, M. Larsson, and S. Pulido, "Affine Volterra processes," Annals of Applied Probability, vol. 29, no. 5, pp. 3155–3200, 2019.
- T. Bollerslev, "Generalized autoregressive conditional heteroskedasticity," Journal of Econo-metrics,

vol. 31, no. 3, pp. 307–327, 1986.
- D. B. Nelson, "Conditional heteroskedasticity in asset returns: A new approach," Econo-metrica, vol. 59, no. 2, pp. 347–370, 1991.
- 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.
- 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.
- F. Corsi, "A simple approximate long-memory model of realized volatility," Journal of Financial Econometrics, vol. 7, no. 2, pp. 174–196, 2009.
- F. X. Diebold and R. S. Mariano, "Comparing predictive accuracy," Journal of Business & Economic Statistics, vol. 13, no. 3, pp. 253–263, 1995.
- P. H. Kupiec, "Techniques for verifying the accuracy of risk measurement models," Journal of Derivatives, vol. 3, no. 2, pp. 73–84, 1995.
- P. F. Christoffersen, "Evaluating interval forecasts," International Economic Review, vol. 39, no. 4, pp. 841–862, 1998.
- 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.
- 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.
- 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.
- K. Christensen, M. V. Siggaard, and B. Veliyev, "A machine learning approach to volatil-ity forecasting," Journal of Financial Econometrics, vol. 21, pp. 1680–1727, 2023. doi: 10.1093/jjfinec/nbac020.
- 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.
- 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.
- 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.
- 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.
- 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.
- 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.
- 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.
- 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.
- J. D. D´ıaz, E. Hansen, and G. Cabrera, "Machine-learning stock market volatility: Pre-dictability, drivers, and economic value," International Review of Financial Analysis, vol. 94, Art. 103286, 2024. doi: 10.1016/j.irfa.2024.103286.
- E. Abi Jaber, "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.
- 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.
- 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.
- R. Kili¸c, "Linear and nonlinear econometric models versus machine-learning models: Ev-idence 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.