Research Square

Preprints are preliminary reports that have not undergone peer review.
They should not be considered conclusive, used to inform clinical practice,
or referenced by the media as validated information.

Signal Persistence in Long-Horizon Risk Projection: Boundary Conditions for the Hybrid GPR-HS Framework

Ujjwala Vadrevu

ujjwala.vadrevu@gmail.com

Independent Researcher

Research Article

Keywords: Gaussian Process Regression, Signal Persistence Horizon, Long-Horizon Forecasting, GARCH, EWMA, Value-at-Risk, CCAR, ICAAP, Length-Scale Governance, Model Risk Management, Historical Crisis Backtesting, Commodity Futures, Emerging Market FX

Posted Date: November 2nd, 2026

DOI: https://doi.org/10.21203/rs.3.rs-11562005/v1

License: This work is licensed under a Creative Commons Attribution 4.0 International License.
Read Full License

Additional Declarations: The authors declare no competing interests.

Signal Persistence in Long-Horizon Risk Projection: Boundary Conditions for the Hybrid GPR-HS Framework

Ujjwala Vadrevu

Replication code: https://colab.research.google.com/drive/1nrlSqmG10DNerNmEqGIh3EB9CcLWIgH9

Abstract

Long-horizon forecasting requires models to retain meaningful predictive differentiation across extended horizons. This study introduces the Signal Persistence Horizon (SPH), a model-agnostic diagnostic measuring how long forecasts retain material differentiation over a specified evaluation horizon, and applies it to nine-quarter volatility projection using Gaussian Process Regression (GPR) and GARCH. Across eight commodity and emerging-market FX series, baseline GPR forecasts rapidly lose differentiation, motivating an investigation of the underlying persistence structure. The Matérn kernel length-scale consistently converges to the imposed optimization boundary across alternative search ranges and specifications, indicating that persistence is not reliably identified by maximum marginal likelihood from daily returns alone. Further experiments separate persistence from signal amplitude and identify distinct failure modes associated with premature forecast convergence and asset-invariant long-horizon variance. A revised specification combining free kernel amplitude with a horizon-aligned, sensitivity-tested length-scale is tested to address these persistence and amplitude failure modes and restores stable, asset-differentiated nine-quarter forecasts. The results suggest that forecast persistence should be treated as an explicit modelling assumption and sensitivity dimension when data do not identify a unique long-horizon persistence structure.

Keywords: Gaussian Process Regression, Signal Persistence Horizon, Long-Horizon Forecasting, GARCH, EWMA, Value-at-Risk, CCAR, ICAAP, Length-Scale Governance, Model Risk Management, Historical Crisis Backtesting, Commodity Futures, Emerging Market FX

1. Introduction

1.1 From Portfolio Validation to Native Long-Horizon Projection

Vadrevu (2026a) established the Hybrid GPR-HS framework and validated it at the portfolio level across seven global equity indices using expanding-window, forward-chaining cross-validation with refitting at each annual split. Vadrevu (2026b) extended the framework to forward-looking stress scenarios using a constructed 252-day shock path combined with Scenario-Averaged Covariance Stabilization (SACS). Both papers reported strong regulator-relevant results, including a 100% portfolio Expected Shortfall (ES) pass rate in the base case and stable Stressed VaR (SVaR) estimates under multiple macroeconomic scenarios.

Neither paper, however, examined whether the framework can serve as a native long-horizon projection engine. Production capital workflows often require forecasts extending to the nine-quarter (9Q) planning horizon used under CCAR (Federal Reserve, 2020), and these projections are frequently generated through asset-class-specific volatility forecasting, scaling, and risk-factor reconstruction procedures. A simpler alternative would be a framework capable of projecting risk directly in return-space without intermediate scaling steps. The study evaluates GPR alongside two widely used volatility forecasting approaches, GARCH (Engle, 1982; Bollerslev, 1986) and EWMA (Longerstaey and Zangari, 1996). Gaussian Process Regression (GPR) provides a flexible non-parametric alternative in which persistence and dependence structures are represented through kernel functions rather than explicitly specified volatility dynamics (Stein, 1999; Rasmussen and Williams, 2005).

This paper investigates that possibility by isolating the univariate GPR component under two conditions not examined previously: (i) forecasting through real historical crisis episodes without scheduled refitting, and (ii) native extrapolation to the CCAR 9Q horizon. The central finding is that long-horizon projection is fundamentally constrained not by the ability of GPR to extrapolate, but by the rapid decay of forecast signal persistence. Understanding the source of this signal loss, and restoring meaningful long-horizon persistence, becomes the primary objective of the study.

1.2 Contributions

This paper makes four contributions:

1. Real historical-crisis reactivity assessment: Using eight documented crisis episodes spanning commodity futures and emerging-market foreign exchange, the paper compares GARCH, EWMA, and univariate GPR-HS under a strict no-refit protocol. The results quantify the cost of GPR's structural non-reactivity and assess a jump-triggered refit policy as a practical mitigation mechanism.
2. Signal Persistence Horizon (SPH): The paper introduces Signal Persistence Horizon (SPH), an operational diagnostic that measures how long a forecasting model's outputs remain materially distinct from their terminal-horizon forecast state over a specified evaluation horizon. SPH provides a model-agnostic measure of forecast persistence and forms the primary diagnostic framework used throughout the study.
3. An empirical explanation for signal collapse in long-horizon projection: Using SPH as the diagnostic lens, the paper demonstrates that the Matérn kernel length-scale is not reliably identifiable through maximum marginal likelihood alone when fit to daily return data. Across

multiple search-bound configurations and asset classes, the optimizer converges to whichever boundary is supplied, indicating that persistence cannot be inferred reliably from the data alone.

4. A corrected specification for native 9Q projection: A revised kernel specification combining a free-amplitude ConstantKernel with a horizon-aligned, sensitivity-tested length-scale is shown to restore forecast persistence and produce stable, asset-differentiated 9Q projections. The resulting framework preserves a key practical advantage of GPR: direct horizon-level forecasting without recursive variance-path generation or square-root-of-time scaling.

2. Methodology

2.1 Instrument Universe

The analysis covers eight instruments spanning two asset classes not examined in the earlier GPR-HS studies: four commodity futures (WTI Crude Oil, Natural Gas, Gold, and Copper) and four emerging-market currency pairs (USD/INR, USD/BRL, USD/TRY, and USD/ZAR). The sample was selected to provide both asset-class diversity and a range of documented historical crisis episodes, enabling the framework to be evaluated under substantially different market conditions and economic drivers.

The instruments differ materially in volatility, market structure, liquidity characteristics, and underlying risk factors. Commodity futures are influenced by factors such as supply-demand imbalances, inventory dynamics, and term-structure effects, while emerging-market foreign exchange reflects macroeconomic, monetary-policy, and sovereign-risk considerations. This diversity allows common forecasting behaviours and artefacts to be distinguished from genuinely asset-specific dynamics.

From a methodological perspective, the objective is not asset-class modelling in the traditional sense, but the evaluation of long-horizon signal persistence within a unified return-space framework. The inclusion of heterogeneous instruments therefore serves as a robustness test: patterns observed consistently across all instruments are more plausibly attributable to the modelling framework and data representation than to the economic characteristics of any single asset class.

2.2 Real Historical Crisis Design

Unlike Vadrevu (2026b), which evaluates stress performance under externally supplied scenario paths through the Scenario-Averaged Covariance Stabilization (SACS) framework, the present study evaluates native forecast persistence and long-horizon extrapolation behaviour. The objective is therefore not scenario consumption, but the assessment of how competing forecasting frameworks behave when confronted with market conditions they have not previously observed.

For this reason, the analysis relies exclusively on real historical crisis episodes rather than synthetic stress trajectories. Using observed market events avoids imposing an arbitrary shock profile and permits forecast persistence to be evaluated against actual market behaviour, consistent with the broader use of historical stress episodes in market-risk assessment (BCBS, 2019). Each instrument is therefore tested on a documented crisis episode drawn from the available sample period. Each instrument's own documented crisis episode is used as the test window (Table 1):

Table 1: Real Historical Crisis Windows Used as Test Data

Instrument	Crisis Episode	Test Window
WTI Crude Oil (CL=F)	April 2020 Negative-Price Crash	2020-04-01 to 2020-04-30

Instrument	Crisis Episode	Test Window
Natural Gas (NG=F)	February 2021 Winter Storm Uri	2021-02-08 to 2021-03-05
Gold (GC=F)	March 2020 COVID Liquidity Crunch	2020-03-01 to 2020-03-31
Copper (HG=F)	March 2020 COVID Demand Shock	2020-02-20 to 2020-03-23
USD/TRY	August 2018 Turkish Lira Crisis	2018-08-01 to 2018-08-31
USD/BRL	March 2020 EM Risk-Off	2020-03-01 to 2020-03-31
USD/ZAR	March 2020 EM Risk-Off	2020-03-01 to 2020-03-31
USD/INR	March 2020 EM Risk-Off	2020-03-01 to 2020-03-31

For each instrument, a lead-in training window immediately preceding the crisis episode is used. Training-window length is treated as a sensitivity dimension, with 63, 90, and 120 trading-day samples corresponding approximately to one, one-and-a-half, and two calendar quarters respectively. Each model is estimated once on the pre-crisis training window and is not re-estimated during the subsequent crisis period.

GARCH(1,1) (Bollerslev, 1986) and EWMA ($\lambda = 0.94$, following the RiskMetrics convention; Longerstaey and Zangari, 1996)) are rolled forward through the crisis window using realized returns as innovations. This constitutes a standard use of both models, since their variance forecasts evolve through recursive updating equations that require only newly observed shocks and do not require model re-estimation.

By contrast, the univariate GPR component of the Hybrid GPR-HS framework is evaluated under pure extrapolation. The model is estimated once on the training window and subsequently asked to generate forecasts at crisis-period dates using the baseline specification established in Vadrevu (2026a): a Matérn 5/2 kernel combined with a White Noise kernel under the Aggressive Noise Initialization (ANI) strategy. No realized crisis-period returns are revealed to the model unless an explicit refit policy is invoked in later experiments.

This asymmetry is intentional and reflects how the competing methodologies process new information. GARCH and EWMA incorporate realized shocks through recursive variance updates, whereas native GPR receives no additional information unless explicitly refit. The purpose of the experiment is therefore not to equalize the methodologies, but to quantify the practical consequences of these differing information-update mechanisms and assess their impact on forecast persistence under previously unseen market conditions.

2.3 Jump-Triggered Refit Protocol

To assess whether GPR's structural non-reactivity can be mitigated, a second protocol evaluates each realized crisis-day return against the rolling standard deviation of the data observed to date, using a jump threshold of $|\ return\ | > 3\sigma$. The 3σ threshold is adopted as a conventional outlier criterion, capturing observations that would be expected to occur infrequently under approximately Gaussian behaviour while avoiding excessive refitting triggered by ordinary day-to-day market noise.

Detected jumps trigger an immediate refit of the GPR model on the expanded history; otherwise, the model remains frozen. The resulting event-triggered refit policy approximates a cost-conscious production implementation while preserving responsiveness to major market dislocations.

2.4 Native 9-Quarter Direct Projection and Signal Persistence Horizon (SPH)

Separately from the crisis-window analysis, each instrument's most recent 756 trading days (approximately three years) preceding a nine-quarter (567 trading day) holdout period are used to estimate

GARCH(1,1) and univariate GPR exactly once. GARCH's native multi-step-ahead forecast (via closed-form recursion, without scaling) is compared against GPR's direct .predict() output at each quarter-end date. This design deliberately avoids the square-root-of-time scaling heuristic commonly used to convert one-day VaR estimates into multi-horizon figures (Jorion, 2006), since such scaling assumes independent and identically distributed returns rather than producing a genuine long-horizon forecast (Hull, 2018). Native projection introduces a question distinct from forecast accuracy: for how long does a forecasting model continue to provide materially differentiated information before converging toward its terminal-horizon forecast state? To evaluate this behaviour, the paper introduces an operational persistence diagnostic termed the Signal Persistence Horizon (SPH).

For a given forecast path, SPH is defined as the earliest forecast horizon at which the forecast enters a specified tolerance band around its terminal forecast value and remains within that band thereafter:

$$SPH_{\epsilon,H} = \min\left\{h: \left|\frac{f_j - f_H}{f_H}\right| < \epsilon \quad \forall j \in \{h, \dots, H\}\right\}$$

where:

f_j= forecast at horizon j, $j \in \{h, \dots, H\}$

f_H= forecast at the terminal evaluation horizon H

ε = materiality threshold

h = candidate persistence horizon

For the empirical analysis presented here, a 5% tolerance relative to the ninth-quarter forecast is adopted as a pragmatic materiality threshold. The tolerance threshold ε is problem-specific and may be determined using the historical forecasting behaviour of the model under consideration. SPH is intended as a model-agnostic measure of forecast persistence, with the threshold serving as an application-specific materiality criterion rather than a universal convergence standard.

Intuitively, SPH measures how long a model continues to "say something new" before forecast convergence occurs. While implemented here using a 5% threshold and a nine-quarter planning horizon, the concept is model-agnostic and may be applied to any forecasting framework exhibiting convergence over a specified evaluation horizon. In the present setting, SPH provides a common basis for comparing forecast persistence across GPR, GARCH, and EWMA while remaining independent of their underlying modelling structures. Importantly, SPH is intended as a diagnostic rather than a prescriptive model-retraining rule. Requiring the forecast to remain within the tolerance band after first entry avoids classifying a temporary threshold crossing as loss of forecast differentiation; it does not require the forecast path itself to be monotonic. Both the terminal evaluation horizon H and materiality threshold ε are application-specific. The resulting SPH may inform model monitoring and refresh decisions, but any decision to retrain or recalibrate the model remains subject to model performance assessment, subject-matter expertise, and the applicable model-governance framework.

2.5 Length-Scale Identifiability Testing

The crisis-reactivity experiments in Sections 2.2 and 2.3 demonstrate that native GPR behaves fundamentally differently from GARCH and EWMA under previously unseen market conditions. While

this establishes the existence of a reactivity gap, it does not explain the rapid loss of forecast persistence subsequently observed in the native nine-quarter projection experiment. Section 2.5 therefore investigates whether the persistence structure learned by the kernel is capable of supporting meaningful long-horizon forecasting.

Within the GPR framework, the Matérn kernel length-scale governs the distance over which observations remain correlated and therefore acts as the model's primary persistence mechanism. If native long-horizon projection is to remain informative across a nine-quarter horizon, the persistence structure learned during model estimation must itself remain meaningful over comparable time scales (Stein, 1999; Rasmussen and Williams, 2005).

Four kernel configurations are evaluated sequentially, each designed to isolate a specific hypothesis regarding the observed loss of forecast persistence:

1. Unconstrained Maximum-Likelihood Search: Length-scale bounds are left intentionally wide, with an effectively unconstrained upper limit of 10,000 trading days. This tests whether maximum marginal likelihood, operating without business constraints, identifies an interior persistence parameter appropriate for long-horizon forecasting.
2. Free-Amplitude Specification: A ConstantKernel component is added while leaving the length-scale search otherwise unconstrained. This tests whether the observed loss of forecast persistence originates primarily from the fixed unit-amplitude assumption embedded in the baseline specification rather than from the persistence parameter itself.
3. Business-Constrained Search: The optimizer is initialized using a variogram-style mean-squared-difference scan across candidate lags ranging from 5 to 252 trading days, with the resulting search restricted to a business-relevant range. This evaluates whether economically plausible constraints can recover stable persistence estimates.
4. Bound-Sensitivity Analysis: The bounded search is repeated using alternative upper bounds of 252, 810, and 10,000 trading days, both with and without the ConstantKernel component. The objective is to determine whether the fitted persistence parameter reflects a genuine optimum in the likelihood surface or merely migrates toward whichever boundary is supplied.

Together, these experiments examine three related questions:

1. Can persistence be inferred directly from return data through likelihood maximization alone?
2. Does forecast persistence depend on signal amplitude as well as length-scale?
3. If persistence cannot be reliably inferred from the data, can it be treated as an explicit, sensitivity-tested modelling assumption for long-horizon risk projection?

3. Results

3.1 Real-Crisis Reactivity: GARCH and EWMA React, GPR Does Not

Table 2 summarizes terminal-day volatility estimates across all three methodologies together with the GARCH persistence parameter ($\alpha + \beta$) and the largest realized single-day return observed during each crisis episode.

Table 2: Terminal-Day Volatility Comparison Across Real Historical Crisis Episodes

Instrument	Crisis Episode	Train Length	α+β (GARCH)	GARCH Terminal Vol	EWMA Terminal Vol	GPR Terminal σ	Worst-Day Return
WTI (CL=F)	Apr 2020 Negative-Price Crash	63	1.0000	18.15%	13.75%	6.84%	-28.18%
		90	1.0000	18.08%	13.75%	5.72%	-28.18%
		120	1.0000	18.08%	13.75%	4.99%	-28.18%
Nat Gas (NG=F)	Feb 2021 Winter Storm Uri	63	0.5138	3.79%	3.15%	3.87%	-4.35%
		90	0.9034	3.79%	3.15%	4.60%	-4.35%
		120	0.8808	4.44%	3.15%	4.46%	-4.35%
Gold (GC=F)	Mar 2020 COVID Liquidity Crunch	63	1.0000	1.43%	2.34%	0.91%	-4.74%
		90	0.7637	1.43%	2.34%	0.85%	-4.74%
		120	0.7505	1.39%	2.34%	0.86%	-4.74%
Copper (HG=F)	Mar 2020 COVID Demand Shock	63	0.5199	1.41%	2.21%	0.98%	-6.93%
		90	0.3611	1.13%	2.21%	0.99%	-6.93%
		120	0.2523	0.87%	2.21%	0.98%	-6.93%
USD/INR	Mar 2020 EM Risk-Off	63	0.1997	0.52%	0.96%	0.33%	-1.87%
		90	0.9933	0.27%	0.96%	0.35%	-1.87%
		120	0.9807	0.27%	0.96%	0.40%	-1.87%
USD/BRL	Mar 2020 EM Risk-Off	63	1.0000	0.73%	1.39%	0.64%	-1.73%
		90	0.9476	0.65%	1.39%	0.68%	-1.73%
		120	0.9676	0.65%	1.39%	0.69%	-1.73%
USD/TRY	Aug 2018 Turkish Lira Crisis	63	0.4373	2.04%	4.28%	1.53%	-8.68%
		90	0.9749	3.81%	4.28%	1.33%	-8.68%
		120	1.0000	4.33%	4.28%	1.18%	-8.68%
USD/ZAR	Mar 2020 EM Risk-Off	63	0.4448	1.89%	1.61%	0.75%	-2.33%
		90	0.4240	1.45%	1.61%	0.77%	-2.33%
		120	0.3549	1.42%	1.61%	0.76%	-2.33%

Three patterns emerge. First, several specifications produce GARCH persistence estimates at or close to the stationarity boundary ($\alpha + \beta \approx 1$)) (Engle, 1982; Bollerslev, 1986). WTI crude oil reaches $\alpha + \beta = 1$ across all three training-window lengths, while boundary solutions also occur for Gold and USD/BRL under the 63-day specification and USD/TRY under the 120-day specification. The variation across training windows further indicates that estimated GARCH persistence can be sensitive to the estimation sample.

Second, GPR consistently records materially lower terminal volatility than both GARCH and EWMA for instruments experiencing the largest realized shocks. WTI crude oil records a terminal GPR volatility of 5.72% compared with 18.08% under GARCH following a realized daily return of −28.18%. Similarly, USD/TRY records 1.33% under GPR compared with 3.81% under GARCH following a realized daily return of −8.68%. Across the crisis episodes studied, native GPR systematically under-reacts whenever realized market conditions diverge substantially from those represented in the pre-crisis training sample.

Third, GPR exhibits materially greater sensitivity to training-window length than either GARCH or EWMA. For several of the most severe crisis episodes, terminal GPR volatility declines as the training

history expands. WTI crude oil decreases from 6.84% to 4.99% as the training window increases from 63 to 120 trading days, while USD/TRY declines from 1.53% to 1.18% over the same range. By contrast, the corresponding GARCH and EWMA estimates remain comparatively stable. This behaviour suggests that the native GPR specification becomes increasingly influenced by the larger body of pre-crisis observations as the training window expands, reducing the relative impact of the subsequent crisis period.

These findings are not unexpected. GARCH and EWMA incorporate newly observed shocks directly through their recursive update mechanisms (Engle, 1982; Bollerslev, 1986; Longerstaey and Zangari, 1996), allowing realized market events to influence forecasts immediately. Under the no-refit protocol adopted here, native GPR receives no new information once estimation is complete and therefore continues to extrapolate from the pre-crisis state of the world. The result is a systematic reactivity gap whenever market conditions change materially.

The significance of this result extends beyond crisis forecasting. The evidence demonstrates that native GPR lacks an intrinsic information-update mechanism comparable to the persistence structures embedded within GARCH and EWMA. This observation motivates the subsequent analysis of forecast persistence: if long-horizon forecasting is to remain informative without frequent retraining, the framework must possess an alternative mechanism capable of carrying meaningful signal across extended horizons.

3.2 Jump-Triggered Refit: Narrows the Gap, Does Not Close It

The structural non-reactivity documented in Section 3.1 raises a natural question: can the gap be reduced through an event-driven refit policy rather than continuous retraining? To evaluate this possibility, the jump-triggered refit protocol described in Section 2.3 was applied across all crisis windows using a $|return| > 3\sigma$ trigger threshold.

Refitting consistently increased terminal GPR volatility estimates relative to the static specification, particularly during episodes characterized by large market dislocations. The effect was most pronounced during the April 2020 WTI negative-price collapse, where the terminal volatility estimate increased from 6.84% to 8.93% under the 63-day training specification, compared with a corresponding GARCH estimate of 18.15%. Similar improvements were observed during the Turkish Lira crisis and the March 2020 commodity-market dislocations.

Despite these improvements, the refit protocol did not fully close the gap relative to GARCH. In the largest crisis episodes, jump-triggered GPR estimates remained materially below the corresponding GARCH forecasts at the end of the stress window. This suggests that the observed reactivity deficit is structural rather than purely operational.

The mechanism can be understood directly from the updating procedure. Each GPR refit recalculates the Aggressive Noise Initialization variance using the entire expanded training history. Newly observed crisis returns are therefore averaged together with a substantially larger body of pre-crisis observations. GARCH's conditional variance recursion behaves differently: the most recent shock enters directly through the ARCH component, giving extreme observations immediate influence regardless of the size of the historical sample (Engle, 1982; Bollerslev, 1986). Event-triggered refitting therefore increases responsiveness but does so through a mechanism that inherently dilutes the impact of large shocks as the training window grows.

Table 3: Impact of Jump-Triggered Refitting on Terminal Volatility

Instrument	Train Length	Refits Triggered	Static GPR Terminal σ	Jump-Triggered GPR Terminal σ	GARCH Terminal Volatility
WTI Crude Oil	63	3	6.84%	8.93%	18.15%
	90	3	5.72%	7.74%	18.08%
	120	4	4.99%	6.87%	18.08%
Natural Gas	63	0	3.87%	3.87%	3.79%
	90	0	4.60%	4.60%	3.79%
	120	0	4.46%	4.46%	4.44%
Gold	63	5	0.91%	1.53%	1.43%
	90	5	0.85%	1.37%	1.43%
	120	5	0.86%	1.28%	1.39%
Copper	63	3	0.98%	1.27%	1.41%
	90	3	0.99%	1.30%	1.13%
	120	3	0.98%	1.17%	0.87%
USD/INR	63	4	0.33%	0.58%	0.52%
	90	4	0.35%	0.54%	0.27%
	120	4	0.40%	0.54%	0.27%
USD/BRL	63	3	0.64%	0.88%	0.73%
	90	3	0.68%	0.86%	0.65%
	120	3	0.69%	0.83%	0.65%
USD/TRY	63	2	1.53%	3.02%	2.04%
	90	3	1.33%	2.62%	3.81%
	120	3	1.18%	2.31%	4.33%
USD/ZAR	63	4	0.75%	1.09%	1.89%
	90	3	0.77%	1.03%	1.45%
	120	4	0.76%	0.98%	1.42%

Two additional observations emerge from Table 3. First, the number of refits required was often substantial relative to the duration of the crisis window. Gold triggered five refits within a 22-trading-day window, while several foreign-exchange crises generated three or four refits over similarly short periods. Even under an event-driven protocol, the resulting retraining cadence approached 15-25% of all crisis-period observations.

Second, the effectiveness of refitting was highly instrument dependent. For Gold, jump-triggered retraining largely eliminated the gap relative to GARCH, whereas for WTI crude oil and USD/TRY, substantial differences remained even after multiple refits. This suggests that limited retraining can improve responsiveness but does not provide a universally reliable substitute for models with explicit recency-weighting mechanisms.

Taken together, the results suggest that jump-triggered refitting represents a practical compromise between complete model stasis and continuous retraining. However, because even relatively frequent refits are unable to consistently replicate the responsiveness of recency-weighted models such as GARCH and EWMA, the question naturally shifts from reactivity to persistence: once generated, how long does

useful forecast information actually survive? The next section addresses this question directly through the Signal Persistence Horizon (SPH) framework.

3.3 Signal Persistence Horizon: Initial Evidence of Rapid Forecast Collapse

The jump-triggered refit results of Section 3.2 demonstrate that responsiveness can be improved through event-driven retraining, but only partially and at the cost of increasingly frequent model updates. This raises a more fundamental question: once generated, how long does useful forecast information survive?

Table 4 reports the Signal Persistence Horizon (SPH) for native nine-quarter projections under the unconstrained length-scale specification (Test 1, Section 2.5). Contrary to the original working hypothesis, forecast persistence proved substantially shorter under GPR than under GARCH. GPR's SPH was one quarter for all eight instruments examined, indicating that the forecasts became materially indistinguishable from their respective terminal-horizon values almost immediately.

By contrast, GARCH's SPH ranged from one to nine quarters depending on the estimated persistence structure. The most extreme example was USD/TRY, where the near-unit persistence estimate ($\alpha + \beta \approx 1$) produced forecast differentiation throughout the full nine-quarter horizon.

Table 4: Signal Persistence Horizon (SPH) Comparison Under Native 9-Quarter Projection

Instrument	GARCH SPH (Quarters)	GPR SPH (Quarters)	Difference (GPR − GARCH)	GARCH Q9 Vol	GPR Q9 Vol
WTI Crude Oil (CL=F)	1	1	0	3.22%	3.98%
Natural Gas (NG=F)	3	1	-2	5.39%	4.39%
Gold (GC=F)	1	1	0	1.16%	1.08%
Copper (HG=F)	1	1	0	1.53%	1.51%
USD/INR	1	1	0	0.40%	0.44%
USD/BRL	2	1	-1	1.21%	1.18%
USD/TRY	9	1	-8	5.32%	1.44%
USD/ZAR	1	1	0	0.99%	1.00%

Averaged across instruments, GPR's SPH was 1.38 quarters shorter than that of GARCH. The result directly contradicted the original expectation that the non-parametric framework would exhibit stronger long-horizon persistence than conventional volatility models.

This finding forms the central empirical puzzle of the paper. If native GPR forecasts lose meaningful differentiation after a single quarter, then direct long-horizon forecasting becomes operationally questionable regardless of short-horizon estimation performance. Understanding the source of this persistence collapse therefore becomes essential.

To investigate the mechanism responsible, the forecast is next decomposed into its mean (μ) and volatility (σ) components before examining the behaviour of the kernel persistence structure itself.

3.4 Length-Scale Non-Identifiability: Four Independent Confirmations

The SPH results presented in Section 3.3 establish that native GPR forecasts lose meaningful forecast differentiation substantially earlier than expected for a nine-quarter projection task. The next question is therefore explanatory rather than descriptive: why does forecast persistence collapse so rapidly?

Within the GPR framework, the Matérn kernel length-scale is the primary mechanism governing persistence. It controls the distance over which observations remain correlated and therefore determines how much information from the training sample can influence long-horizon forecasts (Stein, 1999; Rasmussen and Williams, 2005). If native nine-quarter projection is to remain informative, the optimized length-scale should converge to a stable, economically meaningful memory horizon.

The central diagnostic result of this paper is that no such interior optimum is observed. Across four independent experiments, the fitted length-scale consistently converges to whichever boundary the optimizer is permitted to reach, regardless of where that boundary is placed (Table 5).

Table 5: Evidence of Length-Scale Non-Identifiability Across Alternative Search Bounds

Configuration	Upper Bound Tested	Fitted Length-Scale (all 8 instruments)	ConstantKernel Included?
Test 1: unconstrained search	1×10^4 days	10,000.0 (wall, all instruments)	No
Test 3: bounded, empirically-informed init	252 days	252.0 (wall, all instruments)	No
Test 4a: widened bound	810 days	810.0 (wall, all instruments)	Yes
Test 4b: widened bound, amplitude fixed	810 days	809.9999... (wall, all instruments)	No

Across all four configurations, the fitted length-scale converged to the upper boundary supplied to the optimizer, independent of asset class, search initialization, or inclusion of the ConstantKernel component. This behaviour was observed for all eight instruments examined.

The result is remarkably consistent. Every instrument, spanning both commodity futures and emerging-market foreign exchange, converges to the imposed boundary irrespective of asset class, volatility regime, crisis episode, initialization strategy, or amplitude specification. The optimizer does not discover different persistence horizons for different assets; it simply migrates to the largest value permitted by the search space.

This behaviour is difficult to reconcile with the existence of a well-defined interior maximum in the likelihood surface. Instead, the evidence is consistent with a practical identifiability failure: a situation in which the available return data contains insufficient information for maximum marginal likelihood estimation to uniquely determine a persistence parameter.

One possible explanation lies in the nature of the input data itself. Financial return series are deliberately constructed to suppress trend, remove price-level information, and promote stationarity (Tsay, 2005). The very transformations that make returns suitable for statistical modelling may also attenuate the long-horizon persistence signals required for reliable estimation of a memory parameter. Under this interpretation, the optimizer's repeated migration to the search boundary is not evidence of exceptionally long persistence, but evidence that persistence cannot be inferred robustly from the return series alone.

The implication is significant. Persistence, as represented by the length-scale parameter, does not appear to emerge uniquely from the data under the specification tested here. Instead, it behaves as a governed modelling assumption whose value must be determined through business justification and sensitivity analysis rather than unconstrained likelihood optimization alone.

3.5 Consequence of Non-Identifiability: Two Distinct Failure Modes

The length-scale results of Section 3.4 establish that the Matérn kernel persistence parameter cannot be identified reliably through maximum marginal likelihood. The practical consequence is not merely a

poorly estimated hyperparameter, but a loss of control over forecast persistence itself. Depending on which boundary the optimizer reaches, two distinct failure modes emerge.

At the short-persistence extreme, represented by the unconstrained optimization experiments, forecast persistence collapses almost immediately. This behaviour is directly reflected in the SPH results reported in Section 3.3, where GPR forecasts became materially indistinguishable from their terminal state after a single quarter for all eight instruments. A decomposition of the forecast into its mean (μ) and volatility (σ) components shows that both contribute to this collapse. The posterior mean rapidly reverts toward the zero-mean prior, while the volatility forecast flattens almost immediately. As a result, the forecast ceases to provide meaningful differentiation across the nine-quarter horizon.

At the opposite extreme, persistence can be extended by forcing the optimizer toward materially longer length-scales. However, when this is done under the baseline specification, a different pathology emerges. With the kernel amplitude fixed at its default value, forecast volatility becomes increasingly detached from the underlying asset characteristics and converges toward a common ceiling across all instruments. Rather than producing asset-specific persistence, the model generates a shared artefact.

The mechanism is straightforward. Without a ConstantKernel component, the Matérn kernel's self-covariance $k(x, x)$ is fixed at unity. Consequently, the far-field predictive variance is bounded by approximately $1 + \text{noise_level}$, irrespective of the asset's historical volatility. As the forecast horizon extends sufficiently far beyond the effective memory implied by the fitted length-scale, all instruments approach the same generic variance ceiling.

This behaviour is illustrated in Table 6. Using the same 810-day length-scale bound, the baseline specification produces terminal volatility estimates between approximately 69% and 73% across all eight instruments despite substantial differences in their historical volatility characteristics. The resulting paths exhibit a near-linear convergence toward a common ceiling, a pattern inconsistent with genuine asset-specific dynamics.

Table 6: Terminal (Q9) Volatility With and Without Free Kernel Amplitude, Length-Scale Bound = 810 Days

Instrument	σ at Q9, with ConstantKernel	σ at Q9, without ConstantKernel	Observed behaviour
WTI (CL=F)	4.12%	72.32%	Converges toward common variance ceiling (~70%)
Nat Gas (NG=F)	4.54%	72.52%	Converges toward common variance ceiling (~70%)
Gold (GC=F)	1.11%	70.32%	Converges toward common variance ceiling (~70%)
Copper (HG=F)	1.56%	70.75%	Converges toward common variance ceiling (~70%)
USD/INR	0.45%	69.37%	Converges toward common variance ceiling (~70%)
USD/BRL	1.22%	70.43%	Converges toward common variance ceiling (~70%)
USD/TRY	1.48%	70.67%	Converges toward common variance ceiling (~70%)
USD/ZAR	1.03%	70.22%	Converges toward common variance ceiling (~70%)

The introduction of a ConstantKernel restores a free signal amplitude and therefore an asset-specific variance scale. The effect is immediate. Under identical data, identical training windows, and identical length-scale bounds, the forecast transitions from a shared, implausible volatility explosion to stable and economically differentiated volatility paths. Because the ConstantKernel is the only changed component,

the contrast provides direct evidence that the common-ceiling behaviour originates from the fixed-amplitude specification rather than from the underlying data.

Together, the two failure modes reveal the practical consequences of persistence non-identifiability. At one extreme, persistence collapses and signal disappears almost immediately. At the other, persistence is artificially prolonged but produces asset-invariant forecasts unrelated to observed market behaviour. A viable long-horizon specification therefore requires both sufficient persistence and an asset-specific amplitude structure. The next section demonstrates that combining a free-amplitude ConstantKernel with a horizon-aligned length-scale achieves this balance and restores meaningful forecast differentiation across the full nine-quarter horizon.

3.6 Corrected Specification: Restoring Forecast Persistence

The preceding sections identified two distinct failure modes arising from length-scale non-identifiability. At one extreme, forecast persistence collapsed almost immediately, producing SPH of one quarter for all eight instruments. At the other, longer persistence produced an asset-invariant volatility ceiling unrelated to the underlying data. A viable long-horizon forecasting specification must avoid both pathologies simultaneously.

The corrected specification combines a free-amplitude ConstantKernel with a length-scale bounded at 810 trading days, chosen to be commensurate with the 567-trading-day (nine-quarter) forecasting horizon. Under this specification, forecast volatility remains stable, asset-specific, and differentiated throughout the projection period (Table 7).

Table 7: Corrected Specification (ConstantKernel + 810-Day Length-Scale): Quarterly Volatility Forecasts

Instrument	Q1	Q2	Q3	Q4	Q5	Q6	Q7	Q8	Q9
WTI (CL=F)	4.00%	4.01%	4.02%	4.04%	4.06%	4.08%	4.09%	4.11%	4.12%
Nat Gas (NG=F)	4.41%	4.42%	4.44%	4.46%	4.48%	4.49%	4.51%	4.53%	4.54%
Gold (GC=F)	1.08%	1.08%	1.09%	1.09%	1.10%	1.10%	1.10%	1.11%	1.11%
Copper (HG=F)	1.51%	1.52%	1.52%	1.53%	1.54%	1.54%	1.55%	1.55%	1.56%
USD/INR	0.44%	0.44%	0.44%	0.44%	0.44%	0.45%	0.45%	0.45%	0.45%
USD/BRL	1.18%	1.18%	1.19%	1.19%	1.20%	1.20%	1.21%	1.21%	1.22%
USD/TRY	1.43%	1.44%	1.44%	1.45%	1.46%	1.46%	1.47%	1.47%	1.48%
USD/ZAR	1.00%	1.01%	1.01%	1.01%	1.02%	1.02%	1.03%	1.03%	1.03%

Unlike the fixed-amplitude specification of Table 6, the corrected configuration produces volatility paths that remain both stable and asset-specific throughout the full projection horizon. Forecast volatility increases gradually where persistence is retained, but no common variance ceiling emerges and no evidence of asset-invariant convergence is observed. Commodity futures continue to exhibit higher forecast volatility than emerging-market foreign exchange, while the relative ordering of assets remains broadly consistent with historical risk characteristics.

A decomposition of the forecasts into their mean (μ) and volatility (σ) components confirms that the corrected specification also stabilizes the mean process. For example, WTI's forecast mean evolves from approximately +0.0001 at Q1 to −0.0010 at Q9, remaining several orders of magnitude smaller than the volatility contribution to the resulting VaR estimate. The directional drift observed under the amplitude-constrained specification is no longer present.

Terminal VaR estimates at Q9, computed using the Student's-t specification ($\nu = 5$) adopted in Vadrevu (2026a), are broadly comparable in magnitude to those produced by GARCH. This similarity should not be interpreted as model agreement. Rather, two fundamentally different forecasting mechanisms are converging toward reasonable long-run risk estimates. Under GARCH, the long-run variance is governed by the recursion's unconditional variance term $\omega/(1-\alpha-\beta)$ (Bollerslev, 1986). Under GPR-HS, the long-run volatility floor emerges from the training-window variance embedded within the Aggressive Noise Initialization framework and propagated through the corrected kernel structure.

The central result is therefore not that the corrected specification reproduces GARCH. It is that meaningful forecast persistence can be restored without sacrificing asset differentiation or resorting to square-root-of-time scaling. Native nine-quarter projection under the GPR-HS framework remains viable, provided persistence is treated as an explicit modelling assumption rather than a parameter expected to emerge uniquely from return data.

4. Discussion

4.1 GPR-HS Has Multiple Valid Operating Modes

The results of this paper do not contradict the findings of Vadrevu (2026a) or Vadrevu (2026b); rather, they identify the conditions under which each implementation of the GPR-HS framework is appropriate.

The portfolio-validation framework of Vadrevu (2026a) operates under periodic refitting and one-day-ahead forecasting. The stress-testing framework of Vadrevu (2026b) evaluates scenario consumption under SACS, where the full stressed path is supplied to the model and the primary objective is stable capital estimation under a predefined scenario. Neither design requires prolonged extrapolation into an unobserved future without additional information.

The present paper examines a different problem: native multi-quarter forecasting in the absence of both scheduled refitting and externally supplied scenario paths. Under these conditions, forecast persistence becomes the dominant challenge. The results show that the baseline GPR specification does not naturally retain meaningful signal across the nine-quarter horizon required for long-horizon capital projection, but that this limitation can be mitigated through an explicit persistence specification.

Consequently, the GPR-HS framework should not be viewed as a single model operating identically across all applications. Instead, it supports multiple valid operating modes corresponding to different risk-management objectives:

1. Short-horizon risk estimation, where periodic refitting enables reactive volatility forecasting and the original validation framework applies directly.
2. Scenario-based stress testing, where externally specified paths are consumed using the SACS methodology and forecast persistence is not the primary concern.
3. Native long-horizon projection, where forecast persistence becomes a first-order modelling requirement and the corrected specification developed in this paper is necessary.

This interpretation is consistent with the role played by persistence parameters in other volatility modelling frameworks. GARCH expresses persistence through the $\alpha+\beta$ structure, EWMA through the decay parameter λ, and GPR through the kernel length-scale, although the underlying mathematical mechanisms differ.

Viewed in this way, the contribution of the present paper is not to replace the earlier versions of the framework, but to extend them by defining the conditions under which direct multi-quarter forecasting remains viable.

4.2 Persistence as a Governed Modelling Assumption

The results of Sections 3.3-3.6 suggest that persistence cannot be inferred reliably from daily return data through maximum marginal likelihood optimization alone. Across all eight instruments and multiple alternative search configurations, the Matérn length-scale converged to whichever boundary was supplied rather than to a stable interior solution.

For practical long-horizon forecasting, persistence should therefore be treated as an explicit modelling assumption rather than an optimizer-derived quantity. In the corrected specification, the length-scale is aligned with the forecasting horizon and evaluated through sensitivity analysis rather than selected solely through likelihood maximization.

This treatment is methodologically similar to the use of fixed, disclosed parameters elsewhere in risk modelling. The relevant governance question is not whether a single optimal value exists, but whether forecast behaviour remains stable, interpretable, and fit for purpose across a reasonable range of persistence assumptions. Traditional model monitoring frameworks are predominantly reactive, requiring observable deterioration in model performance metrics before retraining or recalibration actions are triggered. By contrast, the Signal Persistence Horizon (SPH) can be interpreted as a proactive governance diagnostic. Rather than waiting for realized forecasting performance to degrade, SPH provides an estimate of how long forecast outputs remain materially differentiated and therefore potentially informative. In this role, SPH complements conventional monitoring metrics by introducing a forward-looking perspective on model refresh and retraining decisions.

4.3 Limitations

Three limitations should be noted.

First, the nine-quarter direct-projection analysis was conducted using a single training-window specification (756 trading days) for each instrument. While this mirrors common regulatory practice in long-horizon forecasting exercises, the sensitivity of the SPH and length-scale results to alternative training-window lengths was not evaluated and remains an avenue for future investigation.

Second, the jump-triggered refit experiments examined only a single operational policy based on a 3σ trigger and an expanding training window. The results indicate that responsiveness is partially restored under this approach but also suggest a dilution effect as crisis observations are absorbed into an increasingly large historical sample. Alternative refit architectures, particularly fixed-length rolling windows, may produce different outcomes and were not explored in this paper.

Third, the analysis focuses on univariate forecast behaviour and persistence rather than forecast attribution. Although Section 3.5 uses a μ/σ decomposition to diagnose the observed failure modes, the SHARC-based explainability framework introduced in Vadrevu (2026c) was not extended to the long-horizon projection setting examined here. Explaining forecast persistence and persistence collapse through a formal attribution framework remains an open area for future research.

5. Conclusion

This paper examined whether the univariate Gaussian Process Regression (GPR) component of the Hybrid GPR-HS framework can support native long-horizon risk projection without volatility scaling, recursive variance generation, or continuous model refitting. Across eight historical crisis episodes and a nine-quarter projection horizon, three findings emerge.

First, native GPR exhibits a structural reactivity gap relative to GARCH and EWMA because it lacks an intrinsic information-update mechanism. Event-driven refitting improves responsiveness but does not fully replicate the behaviour of recency-weighted parametric volatility models.

Second, the Signal Persistence Horizon (SPH) diagnostic introduced in this paper reveals that native GPR forecasts lose meaningful forecast differentiation substantially earlier than expected under the baseline specification. Investigation of this persistence collapse shows that the Matérn kernel length-scale is not empirically identifiable through maximum marginal likelihood alone, converging consistently to whichever boundary is supplied across multiple experiments, asset classes, and search configurations.

Third, a corrected specification combining a free-amplitude ConstantKernel with a horizon-aligned, sensitivity-tested length-scale restores both forecast persistence and economically plausible, asset-differentiated nine-quarter projections. The evidence suggests that persistence should be treated as a governed modelling assumption rather than a parameter expected to emerge uniquely from return data.

More broadly, the results indicate that reactivity and persistence are distinct modelling objectives. Short-horizon monitoring requires mechanisms that rapidly incorporate new information, while long-horizon projection requires mechanisms that preserve meaningful signal across extended forecast horizons. The Hybrid GPR-HS framework remains operationally viable for native multi-quarter risk projection, provided its persistence assumptions are explicitly stated, justified, and sensitivity-tested.

Acknowledgements

Generative AI disclosure: Microsoft Copilot was used during manuscript preparation for editorial review, language refinement, and feedback on clarity and presentation. All methodological development, model specifications, empirical analyses, interpretation of results, and final manuscript content were independently reviewed and determined by the author, who takes full responsibility for the work.

References

Basel Committee on Banking Supervision [BCBS] (2019) “Minimum capital requirements for market risk.” Bank for International Settlements (BIS).

Bollerslev, T. (1986) “Generalized autoregressive conditional heteroskedasticity,” *Journal of Econometrics*, 31(3), pp. 307–327. Available at: https://doi.org/10.1016/0304-4076(86)90063-1.

Engle, R.F. (1982) “Autoregressive Conditional Heteroscedasticity with Estimates of the Variance of United Kingdom Inflation,” *Econometrica*, 50(4), pp. 987–1007. Available at: https://doi.org/10.2307/1912773.

Federal Reserve (2020) *Comprehensive Capital Analysis and Review 2020 Summary Instructions - March 2020*, *Board of Governors of the Federal Reserve System*. Available at: https://www.federalreserve.gov/publications/comprehensive-capital-analysis-and-review-summary-instructions-2020.htm (Accessed: December 14, 2025).

Hull, J.C. (2018) *Risk Management and Financial Institutions*. John Wiley & Sons.

Jorion, P. (2006) *Value at Risk, 3rd Ed.: The New Benchmark for Managing Financial Risk*. McGraw Hill Professional.

Longerstaey, J. and Zangari, P. (1996) *RiskMetrics Technical Document - Fourth Edition 1996, December*. New York: Morgan Guaranty Trust Company of New York (RiskMetrics). Available at: https://my.liuc.it/MatSup/2006/F85555/rmtd.pdf.

Rasmussen, C.E. and Williams, C.K.I. (2005) *Gaussian Processes for Machine Learning*. Cambridge, Mass: MIT Press.

Stein, M.L. (1999) *Interpolation of Spatial Data: Some Theory for Kriging*. Springer Science & Business Media.

Tsay, R.S. (2005) *Analysis of Financial Time Series*. 1st ed. Wiley (Wiley Series in Probability and Statistics). Available at: https://doi.org/10.1002/0471746193.

Vadrevu, U. (2026a) “A Hybrid Gaussian Process Regression Framework for Stable Volatility-Covariance Estimation: Evidence from Global Equity Indices.” arXiv. Available at: https://doi.org/10.48550/arXiv.2605.17275.

Vadrevu, U. (2026b) “Forward-Looking Stress Testing Under Macro Scenarios: Stable SVaR Estimation Using a Hybrid GPR-HS Framework with SACS.” arXiv. Available at: https://doi.org/10.48550/arXiv.2606.07575.

Vadrevu, U. (2026c) “SHARC: SHAP-Based Interpretability in Machine Learning Risk Models for Regulatory Capital under ICAAP and CCAR.” arXiv. Available at: https://doi.org/10.48550/arXiv.2607.05484.