AIMS Mathematics

https://www.aimspress.com/journal/Math

AIMS Mathematics, 11(9): 29399–29429.
DOI: 10.3934/math.20261168
Received: 01 April 2026
Revised: 01 September 2026
Accepted: 03 September 2026
Published: 11 September 2026

Research article

Explicit error bounds for multivariate nonlinear filtering in stochastic volatility models with spread variance estimation

Jin Zou*

School of Artificial Intelligence, Lishui University, Lishui, Zhejiang, China

* **Correspondence:** Email: mathzj@alumni.sjtu.edu.cn.

Abstract: We considered real-time state estimation for multiple assets in a financial market, where the hidden state consists of logarithmic prices and stochastic volatilities, and spreads were derived as linear combinations of estimated prices. This led to a continuous-time nonlinear filtering problem with state-dependent diffusion. We adopted the improved Yau–Yau algorithm of Yau and Niu, which combines low-discrepancy sampling, high-order kernel approximations, and a local resampling-restart mechanism to mitigate dimension-dependent degradation under the tested low-effective-dimensional configurations. Our main theoretical contribution was an explicit upper bound for the Hardy–Krause variation of the kernel and likelihood function in terms of model parameters; as a model-specific corollary, we derived an explicit error bound for posterior spread-variance estimators. We also provided empirical guidance for selecting the resampling interval. Numerical experiments exhibited a near-N^{-1} randomized quasi-Monte Carlo trend for the tested integrand, supported the qualitative polynomial dependence of the error bound, and demonstrated the importance of correctly specifying the correlation matrix for spread variance estimation. An illustrative case study on crude oil futures demonstrated the practical feasibility.

Keywords: nonlinear filtering; Yau–Yau algorithm; stochastic volatility; quasi-Monte Carlo; pairs trading
Mathematics Subject Classification: 93E11, 91G70, 60G35

1. Introduction

Nonlinear filtering theory, originating from the seminal works of Kalman and Bucy [1], aims to estimate the hidden state of a stochastic dynamical system from noisy observations. In finance, it plays a crucial role in volatility estimation, algorithmic trading, and risk management. However, real-world applications often involve multiple assets, leading to multivariate state spaces that render classical filters such as the extended Kalman filter (EKF), unscented Kalman filter (UKF), and standard particle filters (PF) either inaccurate or computationally prohibitive due to weight degeneracy [2]. A modern perspective on particle filter convergence is provided in [3].

The Yau–Yau filtering framework, introduced in [4], offers an elegant alternative by transforming the problem into solving the Kolmogorov forward equation offline and performing simple multiplicative updates online. This offline-online decomposition significantly reduces online computational cost. Recently, Yau and Niu [5] proposed an improved Yau–Yau algorithm that incorporates quasi-Monte Carlo (QMC) low-discrepancy sampling, high-order multi-scale kernel approximations, and a local resampling-restart mechanism. Their method achieves rigorous error bounds and demonstrates excellent scalability in numerical experiments, mitigating the curse of dimensionality under appropriate low-effective-dimensional conditions.

In this paper, we adapt the improved Yau–Yau algorithm to a realistic financial market model. The state comprises logarithmic prices of multiple assets and their stochastic volatilities, following a Heston-type model with correlated Brownian motions [6]. Spreads, which capture cointegration, are not part of the state but are computed post-filtering as linear combinations of the estimated prices [7].

The challenge of multivariate nonlinear filtering has attracted substantial attention. Standard particle filters [2,3] are asymptotically consistent as the particle number $N \to \infty$, yet their non-asymptotic error typically grows exponentially with the state dimension unless strong mixing conditions—which are absent in persistent stochastic volatility settings—are satisfied [8–10]. Rebeschini and van Handel [11] showed that only under stringent spatial localization assumptions can this curse of dimensionality be circumvented. Ensemble Kalman filters [12, 13] propagate ensembles through nonlinear dynamics but rely on Gaussian approximations, thereby ignoring the inherent non-Gaussianity of state-dependent diffusion and introducing systematic bias in variance estimation. Deep-learning-based parameterizations [14] learn update maps from data but currently lack uniform, non-asymptotic error guarantees across the state space. In contrast, the improved Yau–Yau algorithm [5] provides an explicit offline-online decomposition that converts filtering into a multivariate integration problem, for which quasi-Monte Carlo (QMC) methods achieve favorable convergence rates when the integrand has low effective dimension [15,16]. Recent advances such as multilevel QMC [17] further improve efficiency. The present work contributes to this line by providing explicit, model-dependent upper bounds for the Hardy–Krause variation of the kernel $K_{\Delta t}$ and the likelihood g_y in the stochastic volatility model, revealing precise polynomial dependence on $1/v_{\min}$ and $1/\eta_{\min}$. Multivariate stochastic volatility models have been widely studied for portfolio risk and derivative pricing [18–20], yet they typically impose restrictive correlation structures. This paper extends the standard multivariate Heston model by incorporating a generic positive definite correlation matrix R while preserving the block structure of the diffusion matrix. To the best of our knowledge, the present analysis provides an explicit model-dependent specialization for a generic positive-definite price correlation matrix.

Our main contributions are as follows:

- Explicit upper bounds for the Hardy–Krause variation: We derive explicit upper bounds for the Hardy–Krause variation of the kernel $K_{\Delta t}(x, y)$ and the likelihood function $g_y(x)$ specifically tailored to the stochastic volatility model. By utilizing completing-the-square techniques and Schur complement estimations, we show that these bounds involve specific powers of $\sup_{x\in\Omega} \|a(x)^{-1}\|_2$ and $\det a(x)^{-1/2}$. Consequently, the normalized fixed-target variation $C_A(y_0)$ (Section 5.1) naturally grows as the volatility-of-volatility parameter $\eta_{\min}$ decreases or the minimum volatility $v_{\min}$ decreases.

- Model-specific error bound for spread variance: We derive an absolute error bound for the truncated posterior variance estimator of any spread $s_{ij,t}$ by combining the $L^\infty \to L^\infty$ operator-norm error bound of [5] (Theorem 5.12) with a normalization lemma and a posterior-moment stability lemma ($L^\infty \to L^1 \to$ moment; Lemmas 4.2 and 4.3), neither of which requires a bounded-Lipschitz density bound. The error of the truncated variance estimator is bounded by $O(\sqrt{\Delta t} + D^*(N)/\Delta t)$ up to a logarithmic factor. The bound controls the truncated posterior variance Var_L; extending it to the complete posterior variance on $\mathbb{R}^n$ is stated as an assumption rather than a theorem. Resampling/restart errors are treated as an empirical stabilization mechanism and investigated numerically.

- Empirical verification and correlation analysis: Through numerical experiments, we find empirical consistency with a near-N^{-1} RQMC trend for the tested integrand (consistent with the Koksma–Hlawka inequality), assess the qualitative polynomial dependence of the error bound (Theorem 4.1), and demonstrate the importance of correctly specifying the correlation matrix R for spread variance estimation. Furthermore, we provide heuristic guidance for selecting the restart interval ι in the local resampling mechanism. An illustrative case study on crude oil futures demonstrates the filter's potential for pairs trading strategies. The revised numerical results are generated from a single scripted pipeline. The raw outputs, configuration files, and table-generation scripts are provided to facilitate independent verification.

The paper is organized as follows. Section 2 describes the state-space model and rigorously verifies the regularity assumptions. Section 3 recalls the improved Yau–Yau algorithm and states the generic error bounds. Section 4 presents our original error analysis tailored to the stochastic volatility model, containing complete proofs for the Hardy–Krause variation bounds and spread variance bounds. Section 5 details numerical experiments that find empirical consistency with a near-N^{-1} RQMC trend, assess the error bound, test correlation-structure sensitivity, and discuss financial applications. Section 6 concludes the paper.

2. State-space model for financial markets

We consider a financial market with r distinct assets. All processes are defined on a filtered probability space $(\Omega, \mathcal{F}, \{\mathcal{F}_t\}, \mathbb{P})$.

2.1. Price dynamics with stochastic volatility

Let $P_t = (p_{1,t}, \ldots, p_{r,t})^\top \in \mathbb{R}^r$ denote the logarithmic prices, and let $V_t = (v_{1,t}, \ldots, v_{r,t})^\top \in \mathbb{R}^r$ represent their instantaneous variances. The full hidden state is

$$X_t = \begin{pmatrix} P_t \\ V_t \end{pmatrix} \in \mathbb{R}^n, \quad n = 2r. \tag{2.1}$$

We adopt a Heston-type model with possibly correlated Brownian motions [6]:

$$\mathrm{d}P_t = \mu_P \,\mathrm{d}t + \mathrm{diag}(\sqrt{V_t})\,\mathrm{d}W_t^P, \tag{2.2}$$

$$\mathrm{d}V_t = \kappa_V(\theta_V - V_t)\,\mathrm{d}t + \mathrm{diag}(\eta)\,\mathrm{diag}(\sqrt{V_t})\,\mathrm{d}W_t^V, \tag{2.3}$$

where $\mu_P \in \mathbb{R}^r$ is the drift vector and $\kappa_V = \mathrm{diag}(\kappa_{V,i})$, $\theta_V = \mathrm{diag}(\theta_{V,i})$, and $\eta = \mathrm{diag}(\eta_i)$ are diagonal matrices of mean-reversion rates, long-term variances, and volatility-of-volatility parameters, respectively. The Brownian motions $W_t^P \in \mathbb{R}^r$ and $W_t^V \in \mathbb{R}^r$ may be correlated, with a constant instantaneous correlation matrix $\rho \in [-1, 1]^{r\times r}$ such that $\mathrm{d}\langle W^P, W^V\rangle_t = \rho\,\mathrm{d}t$. To ensure positivity of variances, we assume the Feller condition $2\kappa_{V,i}\theta_{V,i} \geq \eta_i^2$ for each i (the strict form $2\kappa_{V,i}\theta_{V,i} > \eta_i^2$ is imposed in Assumption (A2) below, where it yields the lower-tail exponent $\alpha > 0$).

2.2. Compact stochastic differential equation representation

The system (2.2)–(2.3) can be written compactly as

$$\mathrm{d}X_t = f(X_t)\,\mathrm{d}t + \Sigma(X_t)\,\mathrm{d}W_t, \tag{2.4}$$

where $W_t = (W_t^P, W_t^V)^\top$ is a $2r$-dimensional Brownian motion with covariance structure

$$\mathrm{d}\langle W, W\rangle_t = Q\,\mathrm{d}t, \quad Q = \begin{pmatrix} R & \rho \\ \rho^\top & I_r \end{pmatrix}. \tag{2.5}$$

Here $R \in \mathbb{R}^{r\times r}$ is a generic positive definite correlation matrix governing the instantaneous correlations among the log-price Brownian motions W_t^P, while ρ captures the price-volatility cross-correlations. We assume that $R \succ 0$ and that

$$S_\rho := I_r - \rho^\top R^{-1}\rho \succ 0.$$

By the Schur-complement criterion, this condition is equivalent to the positive definiteness of Q. Equivalently, it can be written as $\|R^{-1/2}\rho\|_2 < 1$. In the special case where $R = I_r$, one recovers the independent price assumption; in general, R can be fully parameterized to reflect realistic cross-asset dependencies. The lower-right block of Q is I_r, which encodes the modeling assumption that the variance-variance Brownian motions W_t^V are uncorrelated across assets; the parameters η_i scale the marginal volatilities of the individual variance processes but do not encode cross-volatility correlation. Extending Q to allow a non-identity vol-vol block $R_V \neq I_r$ is possible at the model-specification level and would replace the Schur complement $S_\rho = I_r - \rho^\top R^{-1}\rho$ by $S'_\rho = R_V - \rho^\top R^{-1}\rho$, introducing an additional spectral-gap parameter $\gamma_{R_V}^{-1} := \lambda_{\min}(R_V)^{-1}$ into the bound of Lemma 2.1; we leave this extension to future work for notational simplicity. The drift and diffusion are

$$f(x) = \begin{pmatrix} \mu_P \\ \kappa_V(\theta_V - v) \end{pmatrix}, \quad \Sigma(x) = \begin{pmatrix} \mathrm{diag}(\sqrt{v}) & 0 \\ 0 & \mathrm{diag}(\eta)\,\mathrm{diag}(\sqrt{v}) \end{pmatrix}, \tag{2.6}$$

with $x = (p, v)$. The diffusion matrix $a(x) = \Sigma(x)Q\Sigma(x)^\top$ is given by

$$a(x) = \begin{pmatrix} \mathrm{diag}(\sqrt{v})\,R\,\mathrm{diag}(\sqrt{v}) & \mathrm{diag}(\sqrt{v})\,\rho\,\mathrm{diag}(\sqrt{v})\,\mathrm{diag}(\eta) \\ \mathrm{diag}(\eta)\,\mathrm{diag}(\sqrt{v})\,\rho^\top\,\mathrm{diag}(\sqrt{v}) & \mathrm{diag}(\eta)^2\,\mathrm{diag}(v) \end{pmatrix}. \tag{2.7}$$

We emphasize that the model is *not* an affine diffusion in the sense of Duffie–Pan–Singleton [21] when the price-price correlation matrix R is non-diagonal: The price covariance block of $a(x)$ contains the entries $R_{ij}\sqrt{v_i v_j}$ for $i \neq j$, which are not affine functions of the variance vector $v = (v_1, \ldots, v_r)$. Consequently the kernel exponent in (3.4) contains the term $\|y - x\|^2_{a(x)^{-1}}/(2\Delta t)$ whose Hessian in x is a state-dependent matrix (not a constant), so the resulting filtering problem is genuinely nonlinear and is treated as such throughout this paper. The only structure retained is the square-root stochastic-volatility form of $\Sigma(x)$ and the uniform ellipticity on Ω, which suffice for the QMC variation bounds of Section 4; we do not rely on any affine structure or affine transform.

2.3. Observation equation

We observe the logarithmic prices contaminated by noise:

$$\mathrm{d}Y_t = HX_t\,\mathrm{d}t + \Gamma_o\,\mathrm{d}W_t^o, \quad Y_0 = 0, \tag{2.8}$$

where $H = [I_r, 0_{r\times r}]$, W_t^o is an r-dimensional standard Brownian motion independent of W_t, and $\Gamma_o \in \mathbb{R}^{r\times r}$ is a constant observation-noise intensity matrix with $B_o := \Gamma_o\Gamma_o^\top > 0$. The case where $\Gamma_o = I_r$ (i.e., $B_o = I_r$) recovers the unit-observation-covariance model; the general formulation allows the observation noise covariance B_o to be a free parameter (the synthetic experiments of Section 5 use $\sigma_o = 0.1$ while the empirical case study does not invoke this linear-observation model and instead uses the return likelihood with noise $\sigma_{\mathrm{ret}} = 0.003$ of Eq (5.4)). The corresponding likelihood is

$$g_{\Delta y}(x) = \exp\Big[\,h(x)^\top B_o^{-1}\,\Delta y - \tfrac{1}{2}\,\Delta t\,h(x)^\top B_o^{-1}\,h(x)\Big],$$

where $h(x) = Hx$. The constants B_0, B_1, B_2 in Theorem 4.1 are defined with $\|B_o^{-1}\|_2$ in place of the implicit unit matrix: $B_0 = R_\Omega\|H\|_2\|B_o^{-1}\|_2\|\Delta y\| + \frac{1}{2}\Delta t R_\Omega^2\|H\|_2^2\|B_o^{-1}\|_2$, $B_1 = \|H\|_2\|B_o^{-1}\|_2\|\Delta y\| + \Delta t\|H\|_2^2\|B_o^{-1}\|_2 R_\Omega$, $B_2 = \Delta t\|H\|_2^2\|B_o^{-1}\|_2$.

The observation model (2.8) should be interpreted as a continuous-time information flow rather than a literal microstructure noise model. The noise W_t^o absorbs model misspecification and observation errors, and its intensity serves as a tuning parameter that controls the filter's trust in the data. In the case study (Section 5.6), we discretize this scheme to daily data, where the daily log-price difference is treated as a single noisy observation.

2.4. Regularity assumptions and domain truncation

To apply the QMC-based nonlinear filtering theory, we impose the following conditions:

Assumption 2.1. (A1) The functions f and Σ are Lipschitz on a bounded domain Ω and satisfy a linear growth condition.

(A2) There exists a bounded domain $\Omega = [-L_p, L_p]^r \times [v_{\min}, v_{\max}]^r$ ($L_p > 0$, $0 < v_{\min} < v_{\max}$) such that, for the finite horizon $T > 0$ and some constants $C, c, c_-, c_+ > 0$, the localization error decomposes as

$$\mathbb{P}(\exists t \in [0, T] : X_t \notin \Omega) \;\le\; E_{\mathrm{loc}} \;=\; E_P(L_p) + E_{V,-}(v_{\min}) + E_{V,+}(v_{\max}),$$

where

$$\begin{aligned} E_P(L_p) &:= \mathbb{P}\Big(\sup_{t\le T}|P_t| > L_p\Big) \;\le\; C\,e^{-c\,L_p^2},\\ E_{V,-}(v_{\min}) &:= \mathbb{P}\Big(\inf_{t\le T}\min_i V_{i,t} < v_{\min}\Big) \;\le\; C\,v_{\min}^{\alpha},\\ E_{V,+}(v_{\max}) &:= \mathbb{P}\Big(\sup_{t\le T}\max_i V_{i,t} > v_{\max}\Big) \;\le\; C\,e^{-c_+\,v_{\max}}, \end{aligned}$$

with $\alpha = \min_i(2\kappa_{V,i}\theta_{V,i}/\eta_i^2) - 1 > 0$ under the *strict* Feller condition $2\kappa_{V,i}\theta_{V,i} > \eta_i^2$; the three bounds are the sub-Gaussian log-price tail, the Cox–Ingersoll–Ross (CIR) power-law lower tail, and the CIR exponential upper tail [21–23]. On Ω, the diffusion matrix $a(x)$ is uniformly positive definite:

$c_1 I \le a(x) \le c_2 I$ for some constants $c_1, c_2 > 0$. We require $R \succ 0$ and $I_r - \rho^\top R^{-1}\rho \succ 0$ so that Q is positive definite, and each variance component satisfies $v_{\min} \le v_i(x) \le v_{\max}$. The strict positivity $v_{\min} > 0$ ensures $\sqrt{v}$ and $1/\sqrt{v}$ are well-defined on Ω, and $v_{\min}$ and $v_{\max}$ are chosen numerically so that $E_{V,-}(v_{\min})$ and $E_{V,+}(v_{\max})$ are negligible.

(A3) The drift f is smooth with bounded derivatives on Ω.

(A4) The observation function $h(x) = Hx$ is linear.

(A5) The initial density π_0 is smooth and supported in Ω.

Remark 2.1 (Integrability and truncation). Truncation to $\Omega = [-L_p, L_p]^r \times [v_{\min}, v_{\max}]^r$ yields truncation error $E_{\text{loc}}(L_p, v_{\min}, v_{\max}) = O(e^{-cL_p^2}) + O(v_{\min}^\alpha) + O(e^{-c_+ v_{\max}})$, whose three terms are $E_P, E_{V,-}, E_{V,+}$ of Assumption (A2), where the price and upper variance tails decay exponentially, and the lower variance tail decays polynomially ($\alpha > 0$) [21, 22]. Choosing $L_p \ge C_0 \sqrt{\log(1/\Delta t)}$, $v_{\max} \ge C_1 \log(1/\Delta t)$, and $v_{\min} \le C_2\, \Delta t^{1/\alpha}$ gives $E_{\text{loc}} \ll \Delta t$. In Theorem 4.4, only the truncated posterior variance is bounded, so E_{loc} does not enter.

Remark 2.2 (Notation: L vs. R). To avoid a notational clash, throughout this paper, $L_p > 0$ denotes the half-width of the log-price box $[-L_p, L_p]^r$ (the truncation radius for the price coordinates), whereas $R \in \mathbb{R}^{r\times r}$ denotes the constant positive-definite price-price correlation matrix introduced in Eq (2.5). The two symbols are unrelated; in particular, the truncation error $E_{\text{loc}}(L_p, v_{\min}, v_{\max})$ depends on L_p (and on $v_{\min}, v_{\max}$) but not on R, while the diffusion matrix $a(x)$ and its inverse depend on R (through Lemma 2.1) but not on L_p.

2.5. *Preliminary explicit estimates for the diffusion matrix*

We now derive explicit bounds for $a(x)$ and its inverse. All estimates hold on the bounded domain Ω. Because $v_{\min} \le v_i \le v_{\max}$, all entries of $a(x)$ are bounded by a constant depending only on $v_{\max}, \eta_{\max}, \|\rho\|_2$, and $\|R\|_2$. Thus, $\|a(x)\|_2 \le C_a$ for all $x \in \Omega$. More importantly, we provide a rigorous bound for the inverse matrix, which is crucial for our main theorems.

Lemma 2.1 (Bounds for the inverse diffusion matrix)**.** *Let $R \succ 0$ and $S_\rho = I_r - \rho^\top R^{-1}\rho \succ 0$ as required by the Schur-complement condition. Define*

$$D_v = \operatorname{diag}(\sqrt{v}), \quad E_\eta = \operatorname{diag}(\eta), \quad \Sigma(x) = \begin{pmatrix} D_v & 0 \\ 0 & E_\eta D_v \end{pmatrix}.$$

Then $a(x) = \Sigma(x) Q \Sigma(x)^\top$ and

$$a(x)^{-1} = \Sigma(x)^{-\top} Q^{-1} \Sigma(x)^{-1}.$$

On the truncated domain Ω,

$$\sup_{x\in\Omega} \|a(x)^{-1}\|_2 \le \frac{\max\{1, \eta_{\min}^{-2}\}}{v_{\min}} \Big[\gamma_R^{-1} + \gamma_\rho^{-1}(1 + \gamma_R^{-1}\|\rho\|_2)^2\Big],$$

where $v_{\min} = \min_i v_i$, $\eta_{\min} = \min_i \eta_i > 0$, $\gamma_R = \lambda_{\min}(R)$, and $\gamma_\rho = \lambda_{\min}(S_\rho)$.

Proof. From the definition $\Sigma(x) = \mathrm{diag}(D_v, E_\eta D_v)$, we have the block-diagonal inverse

$$\Sigma(x)^{-1} = \begin{pmatrix} D_v^{-1} & 0 \\ 0 & D_v^{-1}E_\eta^{-1} \end{pmatrix}, \qquad D_v^{-1} = \mathrm{diag}(1/\sqrt{v_i}),\ E_\eta^{-1} = \mathrm{diag}(1/\eta_i).$$

Its squared spectral norm is the maximum over the diagonal entries:

$$\|\Sigma(x)^{-1}\|_2^2 = \max_i \Big\{ \frac{1}{v_i}, \frac{1}{\eta_i^2 v_i} \Big\} \le \frac{\max\{1, \eta_{\min}^{-2}\}}{v_{\min}}. \tag{1}$$

Next, for the block matrix Q, the Schur complement $S_\rho = I_r - \rho^\top R^{-1}\rho$ is positive definite by assumption. The block-inverse formula gives

$$Q^{-1} = \begin{pmatrix} R^{-1} + R^{-1}\rho S_\rho^{-1}\rho^\top R^{-1} & -R^{-1}\rho S_\rho^{-1} \\ -S_\rho^{-1}\rho^\top R^{-1} & S_\rho^{-1} \end{pmatrix}.$$

To estimate its norm, split $Q^{-1} = M_1 + US_\rho^{-1}V$, where

$$M_1 = \begin{pmatrix} R^{-1} & 0 \\ 0 & 0 \end{pmatrix}, \quad U = \begin{pmatrix} R^{-1}\rho \\ -I_r \end{pmatrix}, \quad V = \begin{pmatrix} \rho^\top R^{-1} & -I_r \end{pmatrix}.$$

Then $\|M_1\|_2 = \|R^{-1}\|_2 = \gamma_R^{-1}$. Moreover,

$$\|U\|_2^2 = \left\|U^\top U\right\|_2 = \left\|\rho^\top R^{-2}\rho + I_r\right\|_2 \le 1 + \gamma_R^{-2}\|\rho\|_2^2,$$

so $\|U\|_2 \le 1 + \gamma_R^{-1}\|\rho\|_2$, and likewise $\|V\|_2 = \|U\|_2$. Since $\|S_\rho^{-1}\|_2 = \gamma_\rho^{-1}$, we obtain

$$\|US_\rho^{-1}V\|_2 \le \gamma_\rho^{-1}\|U\|_2\|V\|_2 \le \gamma_\rho^{-1}(1 + \gamma_R^{-1}\|\rho\|_2)^2.$$

Therefore,

$$\|Q^{-1}\|_2 \le \|M_1\|_2 + \|US_\rho^{-1}V\|_2 \le \gamma_R^{-1} + \gamma_\rho^{-1}(1 + \gamma_R^{-1}\|\rho\|_2)^2. \tag{2}$$

Finally, since $a(x)^{-1} = \Sigma(x)^{-\top}Q^{-1}\Sigma(x)^{-1}$, the sub-multiplicativity of the spectral norm yields

$$\|a(x)^{-1}\|_2 \le \|\Sigma(x)^{-\top}\|_2\ \|Q^{-1}\|_2\ \|\Sigma(x)^{-1}\|_2 = \|\Sigma(x)^{-1}\|_2^2\ \|Q^{-1}\|_2.$$

Inserting the estimates (1) and (2) gives exactly the claimed bound for every $x \in \Omega$, and the supremum over Ω obeys the same inequality. □

Remark 2.3. The bound in Lemma 2.1 is not sharp, as it treats the off-diagonal blocks of $a(x)$ via the diagonal dominance of the maximum norm. A tighter bound can be obtained by applying the Schur complement formula to the block structure of $a(x)$, yielding

$$\|a(x)^{-1}\| \le \frac{C'}{\min_i v_i}$$

with a smaller constant C'. Since the dependence on $1/\min_i v_i$ remains unchanged, the polynomial nature of the complexity estimate in Theorem 4.1 is unaffected. We therefore retain the simpler bound for ease of exposition.

The entries of $a(x)$ do not depend on the price coordinates and are polynomials in $\sqrt{v_i}$ with a total degree of at most two. A first differentiation with respect to a variance coordinate v_i introduces a factor $v_i^{-1/2}$, and each subsequent differentiation introduces at most one further factor v_i^{-1}; on Ω, all such factors are bounded by powers of $v_{\min}^{-1/2} \leq C(1 + \|a^{-1}\|_\infty)$ (indeed $v_{\min}^{-1} \leq \|a^{-1}\|_\infty$, since $\lambda_{\min}(a(x)) \leq v_{\min}$: The price-block diagonal of a contains the entries $v_i R_{ii} = v_i$). We can then prove by induction that for any multi-index α with $|\alpha| \leq m$,

$$\|\partial^\alpha a(x)\|_2 \leq C_m (1 + \|a^{-1}\|_\infty)^m, \tag{2.9}$$

$$\|\partial^\alpha (a^{-1})(x)\|_2 \leq C'_m (1 + \|a^{-1}\|_\infty)^{2|\alpha|+1}, \tag{2.10}$$

where C_m, C'_m are independent of $\|a^{-1}\|_\infty$, and $\|a^{-1}\|_\infty := \sup_{x\in\Omega} \|a(x)^{-1}\|_2$. The estimate (2.9) follows directly from the monomial structure of the entries of a, and the estimate (2.10) follows by induction from the identity $\partial(a^{-1}) = -a^{-1}(\partial a)a^{-1}$, as detailed in the remark below.

Remark 2.4 (Proof of (2.10)). We prove by induction on $m = |\alpha|$. The case where $m = 0$ follows from Lemma 2.1. Assume the bound holds for all multi-indices with $|\beta| \leq m$. For $|\alpha| = m + 1$, write $\partial^\alpha = \partial_j \partial^\beta$ with $|\beta| = m$. Using $\partial_j(a^{-1}) = -a^{-1}(\partial_j a)a^{-1}$ and applying the Leibniz rule to ∂^β of this identity, each term in the expansion consists of products of a^{-1}, $\partial^\gamma a$, and $\partial^\delta(a^{-1})$ with $|\delta| \leq m$. By the induction hypothesis, each $\partial^\delta(a^{-1})$ contributes at most $(1 + \|a^{-1}\|_\infty)^{2|\delta|+1}$, while each $\partial^\gamma a$ is bounded uniformly by (2.9). Summing the exponents from the at most $m + 2$ factors of a^{-1} and their derivatives yields the exponent $2(m + 1) + 1$, completing the induction.

3. Improved Yau–Yau algorithm and generic error bounds

3.1. DMZ equation and splitting

Let $\sigma(x, t)$ be the unnormalized conditional density. It satisfies the Duncan–Mortensen–Zakai (DMZ) equation [24, 25]:

$$\mathrm{d}\sigma = \mathcal{L}^*\sigma \,\mathrm{d}t + \sigma\, h(x)^\top B_o^{-1}\, \mathrm{d}Y_t, \quad \sigma(x, 0) = \pi_0(x), \tag{3.1}$$

where

$$\mathcal{L}^*\sigma = \frac{1}{2}\sum_{i,j} \partial_{ij}(a_{ij}(x)\sigma) - \sum_i \partial_i(f_i(x)\sigma). \tag{3.2}$$

From the DMZ equation, the time-discretized conditional density follows the recursive Bayes update

$$p(x_k \mid Y_{1:k}) = \frac{p(x_k \mid Y_{1:k-1})\, g_{y_k}(x_k)}{\int p(x'_k \mid Y_{1:k-1})\, g_{y_k}(x'_k)\, dx'_k}, \tag{3.3}$$

with observation likelihood $g_{y_k}(x_k) = \exp(h(x_k)^\top B_o^{-1}\, \Delta y_k - \frac{1}{2}\, h(x_k)^\top B_o^{-1}\, h(x_k)\, \Delta t)$.

The improved Yau–Yau scheme updates the density recursively by $\sigma_k = M(\Delta y_k)\exp(A\Delta t)\sigma_{k-1}$, where $A\sigma = \mathcal{L}^*\sigma$ and $M(\Delta y)$ denotes the multiplication operator $[M(\Delta y)\sigma](x) := \exp(h(x)^\top B_o^{-1}\, \Delta y - \frac{1}{2}\, h(x)^\top B_o^{-1}\, h(x)\, \Delta t)\, \sigma(x)$.

3.2. Kernel approximation and discretization via QMC

Following the construction underlying Theorem 4.2 of [5] for the unit diffusion matrix, we use the following state-dependent integral kernel as a *heuristic local Gaussian approximation* to the semigroup $\exp(A\Delta t) = \exp(\mathcal{L}^*\Delta t)$ on the truncated domain Ω (see Remark 3.1):

$$K_{\Delta t}(x,y) = \frac{1}{(2\pi\Delta t)^{n/2}\sqrt{\det a(x)}} \quad \times \exp\biggl(- \frac{\|y-x\|^2_{a(x)^{-1}}}{2\Delta t} - (y-x)^\top a(x)^{-1} f(x) \\ - \Delta t\Bigl(\nabla\cdot f(x) + \frac{1}{2}\|f(x)\|^2_{a(x)^{-1}}\Bigr)\biggr). \tag{3.4}$$

Choosing N low-discrepancy points $\{x_i\}_{i=1}^N \subset \Omega$, the discrete propagator matrix is $F_{\Delta t}(i,j) = K_{\Delta t}(x_j, x_i)\,\Delta x$. The discrete filter state updates via $\sigma^{(k)} = M^{(k)} \odot (F_{\Delta t}^\top \sigma^{(k-1)})$.

Remark 3.1. Theorem 4.2 of [5] is stated for the unit diffusion matrix. The kernel (3.4) is the corresponding state-dependent extension, with the Euclidean metric replaced by $\|\cdot\|_{a(x)^{-1}}$ ($a(x) = \Sigma(x)Q\Sigma(x)^\top$ is well-defined on Ω by Lemma 2.1). It is a *heuristic local Gaussian approximation* to the semigroup of $\mathcal{L}^*$, not its exact heat kernel: The divergence form $\frac{1}{2}\nabla\cdot(a\nabla\cdot)$ and the adjoint form of $\mathcal{L}^*$ differ by $\partial_i a_{ij}$-terms when $a(x)$ varies, so the local consistency is $(L_{\Delta t}u-u)/\Delta t = \mathcal{L}^*u+O(\sqrt{\Delta t})$ [26–28]. The $\sqrt{\Delta t}$ rate in Theorem 4.4 is therefore conditional on this heuristic.

For a function f on $[0,1]^d$, two standard forms of the Hardy–Krause variation are used in the literature:

$$V_{\mathrm{HK}}(f) := \sum_{\emptyset\neq u\subseteq\{1,\dots,d\}} \int_{[0,1]^{|u|}} \left|\frac{\partial^{|u|} f}{\partial x_u}(\boldsymbol{x}_u; \mathbf{1}_{-u})\right| dx_u, \tag{3.5}$$

$$V_{\mathrm{HK}^0}(f) := \sum_{\emptyset\neq u\subseteq\{1,\dots,d\}} \int_{[0,1]^{|u|}} \left|\frac{\partial^{|u|} f}{\partial x_u}(\boldsymbol{x}_u; \mathbf{0}_{-u})\right| dx_u, \tag{3.6}$$

where in both cases the integration is over the *full* projected cube $[0,1]^{|u|}$, and only the value at which the inactive coordinates $\boldsymbol{x}_{-u}$ are frozen differs: $\mathbf{1}_{-u}$ (upper face) for V_{HK} and $\mathbf{0}_{-u}$ (lower face) for V_{HK^0}. The classical (unanchored) Koksma–Hlawka inequality pairs V_{HK} with the classical star discrepancy $D^*(N)$, and the anchored variant of Hickernell [29] pairs V_{HK^0} with the anchored star discrepancy $D^*_{\mathbf{0}}(N)$. Only the two standard forms (3.5)–(3.6) are used in this paper; an "arbitrary-anchor" variant $V^{\boldsymbol{a}}_{\mathrm{HK}}(f) = \sum_{\emptyset\neq u}\int_{[\boldsymbol{a}_u,\mathbf{1}_u]} |\partial_u f(\boldsymbol{x}_u; \boldsymbol{a}_{-u})|\,dx_u$ is non-standard (for $\boldsymbol{a} = \mathbf{1}$, all integration regions degenerate and the variation vanishes identically) and is not adopted. Throughout this paper, we adopt the upper-face anchor $\mathbf{1}$, i.e., the classical V_{HK} of (3.5), which is the convention paired with the classical star discrepancy $D^*(N)$ in the original Koksma–Hlawka inequality.

Since the Koksma–Hlawka variation and discrepancy are defined on the unit cube, we pull the kernel back to $[0,1]^n$ through the affine map $T : \Omega \to [0,1]^n$ defined by $u_i = (p_i + L_p)/(2L_p)$ for $i = 1,\dots,r$ and $u_{r+i} = (v_i - v_{\min})/(v_{\max} - v_{\min})$ for $i = 1,\dots,r$. Denoting $\widetilde{K}_{\Delta t}(u,y) = K_{\Delta t}(T^{-1}(u), y)$, the theoretical quantity analyzed throughout this paper is the classical Hardy–Krause (HK) variation of the pulled-back kernel at the upper-face anchor $\mathbf{1}$:

$$V_{\mathrm{HK}}(\widetilde{K}_{\Delta t}) = \sum_{\emptyset\neq U\subseteq\{1,\dots,n\}} \int_{[0,1]^{|U|}} \left|\partial_U \widetilde{K}_{\Delta t}(u_U; \mathbf{1}_{-U}, y)\right| du_U.$$

With this convention, the inactive normalized coordinates are fixed at one. In physical coordinates, this corresponds to fixing the inactive price coordinates at L_p and the inactive variance coordinates at $v_{\max}$, i.e., the upper-right corner of $\Omega = [-L_p, L_p]^r \times [v_{\min}, v_{\max}]^r$. It matches the convention used in [15, 16] for the upper-face HK variation, and the implementation in Section 5.1 evaluates V_{HK} using this anchor. We do *not* claim that the upper-right anchor is a "conservative worst-case choice"; it is the standard convention.

Lemma 3.1 (Coordinate transform and Jacobian)**.** *For the affine map $T : \Omega \to [0,1]^n$ with inverse $T^{-1}(u) = (2L_p\, u_{1:r} - L_p,\ (v_{\max} - v_{\min})\, u_{r+1:n} + v_{\min})$, the pulled-back kernel $\widetilde{K}_{\Delta t} = K_{\Delta t} \circ T^{-1}$ satisfies*

$$V_{\mathrm{HK}}(\widetilde{K}_{\Delta t}) = \sum_{\emptyset \neq U \subseteq \{1,\dots,n\}} \int_{\Omega_U} \left| \partial_U K_{\Delta t}(x_U; x^+_{-U}, y) \right| dx_U,$$

where $x^+ = T^{-1}(\mathbf{1}) = (L_p\, \mathbf{1}_r,\ v_{\max}\, \mathbf{1}_r)$ is the physical anchor and $\Omega_U = \prod_{i \in U} \Omega_i$ is the projection of Ω onto U ($\Omega_i = [-L_p, L_p]$ for $1 \le i \le r$; $[v_{\min}, v_{\max}]$ for $r < i \le n$).

Proof. Since T is affine and diagonal, T^{-1} is affine with diagonal Jacobian $DT^{-1} = \mathrm{diag}(2L_p\, \mathbf{1}_r,\ (v_{\max} - v_{\min})\, \mathbf{1}_r)$, hence $|\det(DT^{-1})| = (2L_p)^r (v_{\max} - v_{\min})^r > 0$. Because DT^{-1} is diagonal, the chain rule applied to $\widetilde{K}_{\Delta t} = K_{\Delta t} \circ T^{-1}$ produces no cross terms: for every non-empty $U \subseteq \{1, \dots, n\}$,

$$\partial_U \widetilde{K}_{\Delta t}(u, y) = (\partial_U K_{\Delta t})(T^{-1}(u), y) \prod_{i \in U} \frac{dx_i}{du_i} = (\partial_U K_{\Delta t})(T^{-1}(u), y)\, J_U.$$

Fixing $u_{-U} = \mathbf{1}_{-U}$ corresponds, under $x = T^{-1}(u)$, to fixing $x_{-U} = x^+_{-U}$, and the restricted map $u_U \mapsto T^{-1}(u_U; \mathbf{1}_{-U})_U$ sends $[0,1]^{|U|}$ bijectively onto Ω_U, with $du_U = dx_U / J_U$. Substituting into the definition of V_{HK},

$$V_{\mathrm{HK}}(\widetilde{K}_{\Delta t}) = \sum_{\emptyset \neq U \subseteq \{1,\dots,n\}} \int_{[0,1]^{|U|}} \left| \partial_U \widetilde{K}_{\Delta t}(u_U; \mathbf{1}_{-U}, y) \right| du_U = \sum_{\emptyset \neq U \subseteq \{1,\dots,n\}} \int_{\Omega_U} \left| \partial_U K_{\Delta t}(x_U; x^+_{-U}, y) \right| dx_U,$$

which is the claimed identity. □

4. Error analysis for the stochastic volatility model

We now present our theoretical analysis explicitly tailored to the properties of the financial model. By the Koksma–Hlawka inequality and the characterization of the Hardy–Krause variation as a sum of projected L^1 norms of mixed partial derivatives [15,16], bounding the QMC discretization error reduces to controlling V_{HK} in the integrand. All derivative bounds established below hold uniformly on Ω; they therefore bound the anchored variation at every anchor simultaneously and apply, in particular, to the classical Koksma–Hlawka pairing of the star discrepancy $D^*(N)$ with the anchored Hardy–Krause variation.

4.1. Explicit parameter-dependence bounds for the Hardy–Krause variation

Theorem 4.1 (Explicit parameter-dependence bounds for the Hardy–Krause variation)**.** *Let $K_{\Delta t}(x, y)$ be the kernel defined in* (3.4) *and let $g_y(x) = \exp\left(h(x)^\top B_o^{-1}\, \Delta y - \frac{1}{2}\, h(x)^\top B_o^{-1}\, h(x)\, \Delta t\right)$. Then the following* *explicit bounds hold, with all constants independent of* $\|a^{-1}\|_\infty$*:*

$$V_{\mathrm{HK}}(K_{\Delta t}(\cdot, y)) \le C_K(\Delta t)\Big(\sup_{x\in\Omega} \frac{1}{\sqrt{\det a(x)}}\Big)\Big(1 + \|a^{-1}\|_\infty\Big)^{3n}, \tag{4.1}$$

$$V_{\mathrm{HK}}(g_y) \le e^{B_0} \sum_{\emptyset\neq U\subseteq\{1,\dots,n\}} \mathrm{Vol}(\Omega_U)\,|U|!\,(1 + B_1 + B_2)^{|U|}. \tag{4.2}$$

The constant $C_K(\Delta t)$ *is of order* $\Delta t^{-n/2}\sum_{k=1}^{n}\Delta t^{-k}$*. Here* V_{HK} *is the (upper-face) variation at the physical anchor* x^+ *(Lemma 3.1), and* B_0, B_1, B_2 *with* $R_\Omega := \sup_{x\in\Omega}\|x\|_2$ *are as defined after* (2.8)*. The bound displays an exponential envelope* e^{B_0} *in* $\|\Delta y\|$ *rather than a polynomial dependence. This envelope is intrinsic to the Gaussian likelihood.*

Proof. We work in the regime $\Delta t \le 1$, so that negative powers of Δt are ordered by $\Delta t^{-j} \le \Delta t^{-k}$ for $j \le k$. We decompose the kernel as

$$K_{\Delta t}(x, y) = (2\pi\Delta t)^{-n/2} A(x) e^{Q(x,y)},$$

where $A(x) = (\det a(x))^{-1/2}$ and

$$Q(x, y) = -\frac{\|y - x\|^2_{a(x)^{-1}}}{2\Delta t} - (y - x)^\top a(x)^{-1} f(x) - \Delta t\Big(\nabla\cdot f(x) + \frac{1}{2}\|f(x)\|^2_{a(x)^{-1}}\Big)$$

is the exponent appearing in (3.4).

By introducing $z(x, y) = y - x + f(x)\Delta t$ and expanding, one obtains

$$Q(x, y) = -\frac{1}{2\Delta t}\, z(x, y)^\top a(x)^{-1} z(x, y) - \Delta t\,\nabla\cdot f(x),$$

where the term $-\frac{\Delta t}{2}\|f\|^2_{a^{-1}}$ has been absorbed into the quadratic form in z and therefore does not appear separately. This corrects the double-counting of the drift quadratic term.

Since the Hardy–Krause variation is bounded by the sum of projected L^1 norms of mixed partial derivatives [16], it suffices to control $\partial_U K_{\Delta t}$ and $\partial_U g_y$ uniformly in $y \in \Omega$ for every non-empty $U \subseteq \{1, \dots, n\}$.

Estimates for Q (completed-square). Set $\|a^{-1}\|_\infty := \sup_{x\in\Omega}\|a(x)^{-1}\|_2$. For $|U| = 1$, differentiating the corrected Q with respect to x_k gives

$$\partial_k Q = -\frac{1}{2\Delta t}\Big[\langle \partial_k(a^{-1})z, z\rangle + 2\langle a^{-1}z, \partial_k z\rangle\Big] - \Delta t\,\partial_k(\nabla\cdot f),$$

where $\partial_k z = -e_k + \partial_k f\,\Delta t = O(1)$. On the bounded product domain $\Omega = [-L_p, L_p]^r \times [v_{\min}, v_{\max}]^r$, $\|z\|$ is uniformly bounded. Using the estimate (2.10), we obtain $\|\partial_k(a^{-1})\|_2 \le C_1'(1 + \|a^{-1}\|_\infty)^3$, hence every term is bounded by a constant multiple of $(1 + \|a^{-1}\|_\infty)^3\Delta t^{-1}$. Consequently,

$$|\partial_k Q| \le C_{Q,k}(1 + \|a^{-1}\|_\infty)^3\Delta t^{-1}. \tag{4.3}$$

For higher-order derivatives ($|U| \ge 2$), $\partial_U Q$ is obtained by repeatedly differentiating the two groups in the corrected Q. The second group, $-\Delta t\,\nabla\cdot f$, is smooth and its bounds are independent of Δt and

$\|a^{-1}\|_\infty$. In the first group, the global factor $(2\Delta t)^{-1}$ persists, and each differentiation may hit either a^{-1} or z. Applying (2.10) inductively together with the product rule yields, for every non-empty U,

$$|\partial_U Q(x,y)| \le C_{Q,U}(1+\|a^{-1}\|_\infty)^{3|U|}\Delta t^{-1}, \tag{4.4}$$

where $3|U|$ conservatively overestimates the growth rate $2|U|+1$ of $\partial^\alpha(a^{-1})$ (by (2.10)).

Estimates for e^Q and A. By the multivariate Faà di Bruno formula,

$$\partial_U e^{Q(x,y)} = e^{Q(x,y)} \sum_{\pi\in\Pi(U)} \prod_{B\in\pi} \partial_B Q(x,y),$$

where $\Pi(U)$ is the set of all partitions of U. Because $Q(x,y) \le \Delta t\|\nabla\cdot f\|_\infty$, we have $|e^Q| \le e^{\Delta t C_{\nabla f}} =: M_1$. Using (4.4) and noting that $|\pi| \le |U|$, the worst case occurs when π consists of $|U|$ singletons (since $\Delta t \le 1$), giving

$$|\partial_U e^{Q(x,y)}| \le C_U^{(e)}(1+\|a^{-1}\|_\infty)^{3|U|}\Delta t^{-|U|}, \tag{4.5}$$

with $C_U^{(e)} = M_1 \sum_{\pi\in\Pi(U)} \prod_{B\in\pi} C_{Q,B}$ independent of $\|a^{-1}\|_\infty$.

For the amplitude, $A(x) = (\det a(x))^{-1/2}$, since differentiating once yields $\partial_i A = -\frac{1}{2}A\ \mathrm{tr}(a^{-1}\partial_i a)$. Since $\|a^{-1}\|_2 \le (1+\|a^{-1}\|_\infty)$ and $\|\partial_i a\|_2 \le C_1(1+\|a^{-1}\|_\infty)$ by (2.9), we have $|\partial_i A| \le C\,A(1+\|a^{-1}\|_\infty)^2$. Inductively, each further differentiation introduces at most a factor $(1+\|a^{-1}\|_\infty)^2$ via (2.9)–(2.10), giving

$$|\partial_U A(x)| \le C_{A,U}(1+\|a^{-1}\|_\infty)^{2|U|}A(x), \tag{4.6}$$

where $C_{A,U}$ is independent of $\|a^{-1}\|_\infty$. Since $2|U| \le 3|U|$, this contribution is dominated by the e^Q estimate (4.5) in the assembly below.

Assembly of the kernel bound. Applying the Leibniz rule to $K_{\Delta t} = (2\pi\Delta t)^{-n/2}A\,e^Q$ and inserting (4.5)–(4.6), we find for every non-empty U,

$$|\partial_U K_{\Delta t}(x,y)| \le \frac{C_U''}{(2\pi\Delta t)^{n/2}} A(x)(1+\|a^{-1}\|_\infty)^{3|U|}\Delta t^{-|U|},$$

where C_U'' collects the combinatorial constants. Integrating over the projected domain Ω_U (these projected integrals are precisely the terms of the anchored variation, by Lemma 3.1) whose measure satisfies $\mathrm{Vol}(\Omega_U) \le L_\Omega^{|U|}$ with $L_\Omega := \max\{2L_p, v_{\max} - v_{\min}\}$ and summing over all $U \ne \emptyset$ yields

$$V_{\mathrm{HK}}(K_{\Delta t}(\cdot,y)) \le \frac{1}{(2\pi\Delta t)^{n/2}} \sup_{x\in\Omega} A(x) \sum_{k=1}^{n} \tilde{C}_k(1+\|a^{-1}\|_\infty)^{3k}\Delta t^{-k}L_\Omega^k.$$

This establishes (4.1) with $C_K(\Delta t) = \frac{1}{(2\pi)^{n/2}}\sum_{k=1}^n \tilde{C}_k L_\Omega^k \Delta t^{-k-n/2}$.

Bound for g_y. Because $h(x) = Hx$ is linear, the exponent

$$\ell(x) \;=\; h(x)^\top B_o^{-1}\Delta y - \tfrac{1}{2}\|h(x)\|^2_{B_o^{-1}}\Delta t \;=\; (H^\top B_o^{-1}\Delta y)^\top x - \tfrac{\Delta t}{2}\, x^\top H^\top B_o^{-1} H x$$

is a quadratic form in x, not merely affine. Hence its mixed partial derivatives of order $|U| \ge 3$ vanish identically. Set $b := H^\top B_o^{-1}\Delta y \in \mathbb{R}^n$ and $A_H := \Delta t\, H^\top B_o^{-1} H \ge 0$. Then

$$\partial_k \ell = b_k - (A_H x)_k, \qquad \partial_{kl}\ell = -(A_H)_{kl}, \qquad \partial_U \ell = 0 \;\text{ for } |U| \ge 3.$$

On Ω, define the domain radius $R_\Omega := \sup_{x\in\Omega}\|x\|_2$. Using $\|H^\top B_o^{-1}\Delta y\| \le \|H\|_2\|B_o^{-1}\|_2\|\Delta y\|$ and $\|H^\top B_o^{-1}H\| \le \|H\|_2^2\|B_o^{-1}\|_2$, since

$$|\ell(x)| \le \|H\|_2\|B_o^{-1}\|_2\|\Delta y\|\,\|x\|_2 + \tfrac{\Delta t}{2}\|H\|_2^2\|B_o^{-1}\|_2\,\|x\|_2^2 \le R_\Omega\|H\|_2\|B_o^{-1}\|_2\|\Delta y\| + \tfrac{1}{2}\Delta t\, R_\Omega^2\|H\|_2^2\|B_o^{-1}\|_2 =: B_0,$$

we have $|g_y(x)| \le e^{B_0}$. The first and second derivatives of ℓ satisfy

$$|\partial_k\ell(x)| \le \|H\|_2\|B_o^{-1}\|_2\|\Delta y\| + \Delta t\,\|H\|_2^2\|B_o^{-1}\|_2\,R_\Omega =: B_1, \qquad |\partial_{kl}\ell(x)| \le \Delta t\,\|H\|_2^2\|B_o^{-1}\|_2 =: B_2,$$

and $\partial_U\ell = 0$ for $|U| \ge 3$.

By the multivariate Faà di Bruno formula applied to $g_y = e^\ell$,

$$\partial_U g_y(x) \;=\; g_y(x)\sum_{\pi\in\Pi(U)}\prod_{B\in\pi}\partial_B\ell(x).$$

Because all derivatives of ℓ of order at least three vanish, the Faà di Bruno formula contains only singleton and pair blocks. A partition π with p pair blocks and $k-2p$ singleton blocks (where $k=|U|$) contributes $B_1^{k-2p}B_2^p$, and the number of such partitions is $\frac{k!}{2^p\,p!\,(k-2p)!}$. Hence

$$|\partial_U g_y(x)| \;\le\; e^{B_0}\sum_{p=0}^{\lfloor k/2\rfloor}\frac{k!}{2^p\,p!\,(k-2p)!}\,B_1^{k-2p}\,B_2^p.$$

In particular, since the number of partitions of a k-element set into singletons and pairs is at most the Bell number, which does not exceed $k!$, and since $B_1^{k-2p}B_2^p \le (1+B_1+B_2)^{k-p} \le (1+B_1+B_2)^k$ for every $0 \le p \le \lfloor k/2\rfloor$,

$$|\partial_U g_y(x)| \;\le\; e^{B_0}\,k!\,(1+B_1+B_2)^k.$$

Integrating over Ω_U and summing over $U \neq \emptyset$ gives (4.2):

$$V_{\mathrm{HK}}(g_y) \;\le\; e^{B_0}\sum_{\emptyset\neq U\subseteq\{1,\dots,n\}}\mathrm{Vol}(\Omega_U)\,|U|!\,(1+B_1+B_2)^{|U|}.$$

All bounds are independent of $\|a^{-1}\|_\infty$. □

Remark 4.1. The bounds (4.1)–(4.2) are explicit in their parameter dependence but involve combinatorial constants that depend on n and on derivative bounds of f, a over Ω; they are not closed-form. The numerical experiments confirm the qualitative dependence on $1/\eta_{\min}$ and $1/v_{\min}$.

Remark 4.2. The factor $(1+\|a^{-1}\|_\infty)^{3n}$ in the error bound stems from a worst-case isotropic analysis that counts mixed partial derivatives of order at most n, while the conservative power of the inverse-diffusion norm reaches $3n$, and treats every coordinate symmetrically. In the present financial model, the diffusion matrix $a(x)$ inherits the block structure

$$a(x) = \begin{pmatrix} D_v\,R\,D_v & D_v\,\rho\,D_vE_\eta \\ E_\eta D_v\,\rho^\top D_v & E_\eta^2 D_v^2 \end{pmatrix},$$

in which the price coordinates $(p_1,\dots,p_r)$ enter only through D_v, while the volatility coordinates $(v_1,\dots,v_r)$ drive both D_v and E_η. This anisotropy implies that mixed derivatives involving many price coordinates carry far smaller $1/v_{\min}$ and $1/\eta_{\min}$ factors than the isotropic bound suggests. A refined anisotropic treatment that exploits this block structure—for instance, by replacing the uniform exponent $3n$ with a sum $\sum_{k=1}^{n} c_k$, where c_k depends on how many volatility coordinates the k-th derivative touches—is expected to substantially tighten the exponent. We leave the full development of such an anisotropic bound to future work; the present paper content establishes the qualitative polynomial guarantee and documents the gap empirically. Indeed, the numerical experiments in Section 5 show that the observed growth is far milder (for $n = 4$, the empirical exponents are approximately 1.0 in $1/\eta_{\min}$ and 0.86 in $1/v_{\min}$, compared with the theoretical worst-case predictions 24 and 12). For fixed state dimension n, the bound is polynomial in the inverse non-degeneracy parameters. The polynomial degree and the combinatorial constants grow with n, and the theorem therefore does not establish polynomial complexity in the state dimension.

4.2. *Error bounds for spread variance*

The proof of Theorem 4.4 propagates the assumed pathwise baseline L^∞ density error (Assumption 4.3) on Ω, $\|\sigma_k - \widehat{\sigma}_k\|_\infty \le C_\sigma(\sqrt{\Delta t} + D^*(N)/\Delta t)$, through a normalization lemma ($L^\infty \to L^1$) and a moment-stability lemma ($L^1 \to$ moment), avoiding any bounded-Lipschitz density bound. Two further assumptions are needed: A lower bound on the normalizing mass (Assumption 4.1) and a bounded Hardy–Krause variation of the filtering density (Assumption 4.2), the latter closing the Koksma–Hlawka product-variation gap in [5, Theorem 5.7], where merely bounded φ is not a guaranteed finite variation.

Assumption 4.1 (Normalizing mass lower bound)**.** For a fixed observation path $Y_{0:T}$, there exists $z_0 = z_0(Y_{0:T}) > 0$ such that, for all k and all sufficiently small $\Delta t, N^{-1}$, the unnormalized truncated mass $Z_k = \int_\Omega \sigma_k \, \mathrm{d}x$ and its QMC estimate $\widehat{Z}_k = \int_\Omega \widehat{\sigma}_k \, \mathrm{d}x$ satisfy $Z_k, \widehat{Z}_k \ge z_0$. Path-dependence is unavoidable (the Gaussian likelihood can be arbitrarily small for extreme $\|\Delta y_k\|$), and the resulting bound is therefore *pathwise conditional*.

Assumption 4.2 (Bounded variation of the filtering density)**.** For all k and paths $Y_{0:k}$, the unnormalized truncated density σ_k has a bounded, anchored (upper-face) Hardy–Krause variation

$$\sup_{0\le k\le K} V_{\mathrm{HK}}(\sigma_k) \;\le\; V_{\max} \;<\; \infty, \qquad V_{\mathrm{HK}}(f) \;:=\; \sum_{\emptyset\neq U\subseteq\{1,\dots,n\}} \int_{\Omega_U} \left|\partial^U f(\boldsymbol{x}_U, \boldsymbol{a}_{\bar{U}})\right| \mathrm{d}\boldsymbol{x}_U,$$

a norm under which bounded-HK functions form an algebra: the complete Leibniz rule $\partial^U(fg) = \sum_{V\subseteq U} \partial^V f \, \partial^{U\setminus V} g$ retains every mixed cross term, so $V_{\mathrm{HK}}(f), V_{\mathrm{HK}}(g) < \infty$ implies $V_{\mathrm{HK}}(fg) < \infty$ (the $V = \emptyset$ term is controlled by $\|f\|_\infty, \|g\|_\infty$). The same bound holds for the QMC estimate $\widehat{\sigma}_k$.

Assumption 4.3 (Pathwise baseline density error)**.** For a fixed observation path $Y_{0:k}$ and all sufficiently small $\Delta t, N^{-1}$, the theoretical continuous-density scheme satisfies the baseline error

$$\|\sigma_k - \widehat{\sigma}_k\|_\infty \;\le\; C_\sigma\Big(\sqrt{\Delta t} + \frac{D^*(N)}{\Delta t}\Big),$$

with C_σ independent of $\Delta t, N$ (depending on the model parameters and on the HK-variation bound of Theorem 4.1). Establishing this bound for the state-dependent heuristic kernel—including boundary localization, stability, and error accumulation—is the contribution of [5, Theorem 5.12] under its own hypotheses, and is *assumed* here rather than re-derived.

Assumption 4.4 (Posterior-tail localization estimate)**.** For the spread $s_t = c^\top X_t + c_0$, there exists a deterministic constant $E_{\text{tail}} > 0$ such that, along the observation path $Y_{0:t}$,

$$\mathbb{E}\Big[(s_t - \mathbb{E}s_t)^2\, \mathbf{1}_{\Omega\setminus\Omega_L}\,\Big|\, Y_{0:t}\Big] \;\le\; E_{\text{tail}},$$

where $\Omega_L := \{|p_i| \le L_p,\ v_{\min} \le v_i \le v_{\max}\}$. This assumption is required only to bridge the truncated variance Var_L of Theorem 4.4 to the complete posterior variance on $\mathbb{R}^n$. Therefore, it is stated as an assumption rather than a theorem (cf. Remark 2.1).

Remark 4.3. Assumption 4.2 is needed because [5, Theorem 5.7] applies Koksma–Hlawka to the product $K_{\Delta t}(\cdot, y)\varphi(\cdot)$ with $\|\varphi\|_\infty \le 1$. The product-variation inequality requires $V_{\text{HK}}(\varphi) < \infty$, which a merely bounded φ need not satisfy. Here $\varphi = \sigma_k$ inherits this regularity from the smooth initial density, the smooth Heston coefficients, and the bounded likelihood g_y. $V_{\max}$ depends on model parameters but not on $\Delta t, N$.

Lemma 4.2 (From unnormalized L^∞ error to normalized L^1 error)**.** *Let $\sigma, \widehat{\sigma} \in L^\infty(\Omega)$ be two non-negative unnormalized densities on the bounded domain Ω with $|\Omega| := \mathrm{Vol}(\Omega)$. Assume $\|\sigma - \widehat{\sigma}\|_\infty \le \varepsilon$, $Z := \int_\Omega \sigma\, \mathrm{d}x \ge z_0 > 0$, and $\widehat{Z} := \int_\Omega \widehat{\sigma}\, \mathrm{d}x \ge z_0 > 0$. Define the normalized densities $p = \sigma/Z$ and $\widehat{p} = \widehat{\sigma}/\widehat{Z}$. Then*

$$\|p - \widehat{p}\|_1 \;\le\; \frac{2\,|\Omega|}{z_0}\,\varepsilon.$$

Proof. By the triangle inequality and $\int_\Omega \sigma\, \mathrm{d}x = Z$,

$$\begin{aligned}
\|p - \widehat{p}\|_1 &= \int_\Omega \Big|\frac{\sigma}{Z} - \frac{\widehat{\sigma}}{\widehat{Z}}\Big|\, \mathrm{d}x = \frac{1}{Z\widehat{Z}} \int_\Omega \big|\sigma\widehat{Z} - \widehat{\sigma}Z\big|\, \mathrm{d}x = \frac{1}{Z\widehat{Z}} \int_\Omega \big|\sigma(\widehat{Z} - Z) + Z(\sigma - \widehat{\sigma})\big|\, \mathrm{d}x \\
&\le \frac{1}{Z\widehat{Z}}\Big(|\widehat{Z} - Z| \cdot Z + Z \cdot \|\sigma - \widehat{\sigma}\|_1\Big) \le \frac{1}{Z\widehat{Z}}\Big(|\Omega|\varepsilon \cdot Z + Z \cdot |\Omega|\varepsilon\Big) = \frac{2\,|\Omega|\,\varepsilon}{\widehat{Z}} \le \frac{2\,|\Omega|\,\varepsilon}{z_0},
\end{aligned}$$

where we used $|\widehat{Z} - Z| \le \int_\Omega |\widehat{\sigma} - \sigma|\, \mathrm{d}x \le |\Omega|\,\varepsilon$ and $\|\sigma - \widehat{\sigma}\|_1 \le |\Omega|\,\varepsilon$, together with $\widehat{Z} \ge z_0$. □

Lemma 4.3 (From L^1 density error to posterior moment error)**.** *Let $p, \widehat{p}$ be two probability densities on Ω and let $\varphi \in L^\infty(\Omega)$. Then*

$$\Big|\mathbb{E}_p[\varphi] - \mathbb{E}_{\widehat{p}}[\varphi]\Big| \;\le\; \|\varphi\|_\infty\, \|p - \widehat{p}\|_1.$$

In particular, for a general linear spread $s(x) = c^\top x + c_0$ on Ω, we set $M_s := \|s\|_\infty$ (recovering $M_s \le 2L_p$ for the pair spread $s_{ij} = x_i - x_j$)

$$\Big|\mathbb{E}_p[s] - \mathbb{E}_{\widehat{p}}[s]\Big| \le M_s\, \|p - \widehat{p}\|_1, \qquad \Big|\mathbb{E}_p[s^2] - \mathbb{E}_{\widehat{p}}[s^2]\Big| \le M_s^2\, \|p - \widehat{p}\|_1,$$

and consequently

$$\Big|\mathrm{Var}_p(s) - \mathrm{Var}_{\widehat{p}}(s)\Big| \;\le\; 3\, M_s^2\, \|p - \widehat{p}\|_1.$$

Proof. The first inequality is the duality between L^1 and L^∞. For s^2, we use $\|s^2\|_\infty \le M_s^2$. For the variance we write $\mathrm{Var}_p(s) = \mathbb{E}_p[s^2] - (\mathbb{E}_p[s])^2$ and use

$$\Big|(\mathbb{E}_p[s])^2 - (\mathbb{E}_{\widehat{p}}[s])^2\Big| = \Big|\mathbb{E}_p[s] - \mathbb{E}_{\widehat{p}}[s]\Big| \cdot \Big|\mathbb{E}_p[s] + \mathbb{E}_{\widehat{p}}[s]\Big| \le 2M_s \Big|\mathbb{E}_p[s] - \mathbb{E}_{\widehat{p}}[s]\Big| \le 2M_s^2\, \|p - \widehat{p}\|_1,$$

since $|\mathbb{E}_p[s]|, |\mathbb{E}_{\widehat{p}}[s]| \le M_s$. Combining with the s^2 bound yields $|\mathrm{Var}_p(s) - \mathrm{Var}_{\widehat{p}}(s)| \le M_s^2\|p - \widehat{p}\|_1 + 2M_s^2\|p - \widehat{p}\|_1 = 3M_s^2\|p - \widehat{p}\|_1$. □

Theorem 4.4 (Error bound for truncated posterior spread variance)**.** *Let $s_t = c^\top X_t + c_0$ be a general linear spread of the log-price coordinates (covering both the pair spread $s_{ij,t} = p_{i,t} - p_{j,t}$ and the cointegration spread $s_t = \log P_{i,t} - \hat{\alpha} - \hat{\beta}\log P_{j,t}$), $\widehat{\mathrm{Var}}(s_t \mid Y_{0:t})$ the QMC posterior-variance estimate, and $\mathrm{Var}_L(s_t \mid Y_{0:t})$ the posterior variance under the truncated posterior $p_L = \sigma_L/Z_L$ on Ω. Under Assumptions 2.1, 4.1, and 4.2, and the pathwise baseline error of Assumption 4.3, with the posterior-tail localization estimate of Assumption 4.4 required to bridge Var_L to the complete posterior variance, there is a constant C_{var}, independent of $\Delta t, N$ but depending on the model parameters ($|\Omega|$, z_0^{-1}, M_s, C_σ), such that*

$$\left|\mathrm{Var}_L(s_t \mid Y_{0:t}) - \widehat{\mathrm{Var}}(s_t \mid Y_{0:t})\right| \le C_{\mathrm{var}}\Big(\sqrt{\Delta t} + \frac{D^*(N)}{\Delta t}\Big). \tag{4.7}$$

Proof. Let $\sigma_k, \widehat{\sigma}_k$ be the unnormalized exact and QMC densities at τ_k on $\Omega = [-L_p, L_p]^r \times [v_{\min}, v_{\max}]^r$. By the pathwise baseline error of Assumption 4.3,

$$\|\sigma_k - \widehat{\sigma}_k\|_\infty \;\le\; C_\sigma\Big(\sqrt{\Delta t} + \frac{D^*(N)}{\Delta t}\Big), \tag{4.8}$$

where C_σ depends on $\gamma_R^{-1}, \gamma_\rho^{-1}, V_{\max}$ but not $\Delta t, N$. The $\sqrt{\Delta t}$ (not Δt) term reflects the first-order consistency of the state-dependent kernel (Remark 3.1).

Set $\varepsilon := C_\sigma(\sqrt{\Delta t} + D^*(N)/\Delta t)$. By Assumption 4.1, $Z_k, \widehat{Z}_k \ge z_0$, so $p_k = \sigma_k/Z_k$ and $\widehat{p}_k = \widehat{\sigma}_k/\widehat{Z}_k$ are well-defined, and Lemma 4.2 gives

$$\|p_k - \widehat{p}_k\|_1 \;\le\; \frac{2\,|\Omega|}{z_0}\,\varepsilon. \tag{4.9}$$

For the linear spread s with $M_s = \|s\|_\infty$, Lemma 4.3 (with $\varphi \in \{s, s^2\}$) gives

$$\left|\mathbb{E}_{p_k}[s] - \widehat{\mathbb{E}}[s]\right| \le \frac{2M_s\,|\Omega|}{z_0}\,\varepsilon, \qquad \left|\mathbb{E}_{p_k}[s^2] - \widehat{\mathbb{E}}[s^2]\right| \le \frac{2M_s^2\,|\Omega|}{z_0}\,\varepsilon,$$

and, by its variance-stability bound,

$$\left|\mathrm{Var}_{p_k}(s) - \widehat{\mathrm{Var}}(s)\right| \;\le\; \frac{6\,M_s^2\,|\Omega|}{z_0}\,\varepsilon. \tag{4.10}$$

Substituting ε and identifying $p_k = p_L$ yields the claim with $C_{\mathrm{var}} := \frac{6\,M_s^2\,|\Omega|}{z_0}C_\sigma$ (for the pair spread, $M_s \le 2L_p$ gives $C_{\mathrm{var}} \le \frac{24\,L_p^2\,|\Omega|}{z_0}C_\sigma$). □

Remark 4.4 (Theoretical algorithm vs. empirical implementation). Theorem 4.4 is concerned with the theoretical algorithm with a continuous (grid-interpolated) density on Ω, for which $V_{\mathrm{HK}}(\sigma_k)$ is well-defined. The empirical particle implementation (Section 5) is a discrete measure without a classical density, so its HK variation is defined only after interpolation. Its agreement with the bound is therefore studied numerically (Section 5.4) rather than asserted by the theorem.

5. Numerical experiments

We conduct extensive numerical experiments to examine the qualitative polynomial dependence predicted by Theorem 4.1 and to assess the algorithm's practical performance.

5.1. Experimental methodology and parameter settings

All numerical experiments were conducted using the improved Yau–Yau algorithm with Sobol low-discrepancy sequences. The mixed derivatives required for the Hardy–Krause variation $V_{\mathrm{HK}}(K_{\Delta t})$ were computed using automatic differentiation (AD) via JAX `jacfwd` (forward-mode AD). All computations used IEEE-754 `float64` via `jax.config.update ("jax_enable_x64", True)`. The relative difference in C_A between a genuine `float32` pipeline (JAX with `x64` disabled) and the `float64` pipeline is below 7×10^{-7} (Supplementary Table S5). A step-size sensitivity analysis confirmed that AD eliminates finite-difference truncation error; see the full results reported in Supplementary Data S3 and S4.

Analytic unit tests for the JAX AD. The `jacfwd` mixed-derivative indices were validated against three closed-form tests (a multivariate polynomial, a Gaussian kernel, and a one-dimensional $\sin(x)e^x$), all agreeing with the analytic derivatives to machine precision (relative error below 10^{-14}).

We adopt the upper-face anchor $\boldsymbol{a} = \mathbf{1}$ (physical upper-right corner) [15, 16] as a deterministic convention, where a random-anchor check (Table S6) leaves the qualitative scaling unchanged.

The L^1 norm over $\Omega_{\mathcal{U}}$ used tiered Gauss–Legendre quadrature (20 points per dimension for $|U| \leq 2$, 10 for $|U| = 3$, 5 for $|U| = 4$), the 4th-order derivatives contribute only $\sim 4\%$ of V_{HK}, and comparing 20/20/10/5 with 30/30/15/10 changes V_{HK} by $< 0.15\%$. Quadrature/domain sensitivity is reported in Section 5.7 (Table S7). The variation was evaluated at a fixed domain-center target $y_0 = (0, 0, v_{\min} + R_v, v_{\min} + R_v)$, not the uniform-in-$y$ bound $\sup_y V_{\mathrm{HK}}(K_{\Delta t}(\cdot, y))$. To isolate the dependence on $\|a^{-1}\|_\infty$, we define the normalized fixed-target variation as:

$$C_A(y_0) := \frac{V_{\mathrm{HK}}(K_{\Delta t}(\cdot, y_0))}{\sup_{x\in\Omega}(\det a(x))^{-1/2}}. \tag{5.1}$$

Synthetic data were generated by the Euler–Maruyama scheme at the same time step as the filter, i.e., single-step simulation without sub-stepping to a finer reference grid, and observations follow (2.8) with $\sigma_o = 0.1$. This price-level observation noise σ_o is larger than the return-likelihood noise $\sigma_{\mathrm{ret}} = 0.003$ used in the empirical study (Section 5.6). The restart sweep (Table 3) used $\Delta t = 0.05$, $N_R = 256$, 100 replications, and the correlation experiment (Table 5) used $\Delta t = 0.1$, $N_R = 512$, 1000 replications. The C_A values in Tables 1 and 2 are deterministic (no seed averaging). In the implemented numerical filters, the prediction step was evaluated by applying the state-dependent kernel matrix on Sobol support points; no finite-difference or Crank–Nicolson partial differential equation (PDE) solver was used in the reported experiments. Confidence intervals for the spread estimates were derived from the posterior standard deviation. The QMC convergence experiment (Section 5.4) directly measures the kernel integration error, isolating it from filtering recursion errors.

5.2. Empirical consistency check for qualitative polynomial dependence

We fix $r = 2$ ($n = 4$), $\Delta t = 0.01$, with model parameters $\mu_P = 0$, $\kappa_V = \mathrm{diag}(2, 2)$, $\theta_V = \mathrm{diag}(0.25, 0.25)$, $\rho_{12} = 0.3$ (the other correlations are zero). The variation is computed for $K_{\Delta t}(\cdot, y)$ with y fixed at the domain center, on $\Omega = [-R_p, R_p]^2 \times [v_{\min}, v_{\min} + 2R_v]^2$ with $R_p = 0.02$, $R_v = 0.01$ (domain sensitivity is shown in Table S7).

Experiment 1: Varying $\eta_{\min}$

We vary $\eta_1 = \eta_2 = \eta_{\min}$ from 0.1 to 0.5 at fixed $v_{\min} = 0.2$. Table 1 reports C_A (deterministic, upper-face anchor $a = 1$). A log-log regression of C_A versus $1/\eta_{\min}$ gives $\widehat{\beta_\eta} = 1.026 \pm 0.044$ ($R^2 = 0.9874$; ordinary least squares (OLS) slope SE).

Table 1. Normalized fixed-target variation $C_A(y_0)$ for varying $\eta_{\min}$ ($v_{\min} = 0.2$), computed via automatic differentiation (JAX `jacfwd`) with the upper-face anchor $a = 1$; values are deterministic.

$\eta_{\min}$	$1/\eta_{\min}$	C_A	$\log_{10} C_A$
0.10	10.00	1.491×10^3	3.17
0.15	6.67	1.187×10^3	3.07
0.20	5.00	891.15	2.95
0.25	4.00	688.74	2.84
0.30	3.33	553.17	2.74
0.35	2.86	460.57	2.66
0.40	2.50	395.19	2.60
0.45	2.22	347.18	2.54
0.50	2.00	310.42	2.49

Experiment 2: Varying $v_{\min}$

We vary $v_{\min}$ from 0.1 to 0.5 at fixed $\eta = \mathrm{diag}(0.3, 0.3)$. Table 2 reports C_A (with the same deterministic upper-face anchor as Table 1). The empirical scaling is $\widehat{\beta_v} = 0.856 \pm 0.014$ ($R^2 = 0.9982$; OLS slope SE). This is a *joint* sensitivity: Varying $v_{\min}$ shifts the domain lower bound and the fixed target y_0 simultaneously, so $\widehat{\beta_v}$ captures their combined effect rather than $v_{\min}$ alone.

Table 2. Normalized fixed-target variation $C_A(y_0)$ for varying $v_{\min}$ ($\eta = 0.3$), computed via automatic differentiation (JAX `jacfwd`) with the upper-face anchor $a = 1$; values are deterministic.

$v_{\min}$	$1/v_{\min}$	C_A	$\log_{10} C_A$
0.10	10.00	942.63	2.97
0.15	6.67	705.27	2.85
0.20	5.00	553.17	2.74
0.25	4.00	452.60	2.66
0.30	3.33	383.44	2.58
0.35	2.86	331.78	2.52
0.40	2.50	294.16	2.47
0.45	2.22	266.52	2.43
0.50	2.00	245.32	2.39

Interpretation of the empirical exponents

For $n = 4$, Theorem 4.1 predicts an isotropic worst-case dependence $\propto (1 + \|a^{-1}\|_\infty)^{12}$, but the observed scaling in $1/\eta_{\min}$ and $1/\nu_{\min}$ is far milder (Tables 1 and 2), reflecting Heston anisotropy. The fitted slopes are descriptive and consistent only with the *qualitative* polynomial guarantee, not the quantitative exponent. Slope SEs are from OLS residuals (7 d.f.); the reported intervals reflect only this residual variance, not quadrature or sampling error (a bootstrap treatment is left to future work).

5.3. Empirical analysis of the restart interval

We use a synthetic dataset with $r = 5$ assets ($n = 10$), $T = 1.5$, $\Delta t = 0.05$ (30 steps), and local sample size $N_R = 256 = 2^8$ (with a power of two preserving the Sobol digital balance). Model parameters are $\mu_P = 0$, $\kappa_V = 2I_5$, $\theta_V = 0.25I_5$, $\eta = 0.3I_5$, $R_{ij} = 0.2$ for $i \neq j$, $\rho = 0$, where observations follow (2.8) with $\sigma_o = 0.1$. Each configuration is averaged over 100 runs.

We vary the restart interval ι over $\{1, 2, 3, 5, 8, 10, 12, 15, 20\}$ and compute the root mean square error (RMSE) of the estimated mean price trajectory:

$$\mathrm{RMSE} = \sqrt{\frac{1}{r\,T/\Delta t} \sum_{k=1}^{T/\Delta t} \sum_{i=1}^{r} (\hat{p}_{i,k} - p^{\mathrm{ref}}_{i,k})^2}, \tag{5.2}$$

where $p^{\mathrm{ref}}_{i,k}$ denotes the true simulated price of asset i at step k from the data-generating path.

Table 3 reports RMSE (mean ± std over 100 runs) together with four degeneracy diagnostics (mean ESS $\overline{\mathrm{ESS}}$, minimum per-step ESS $\min_t \mathrm{ESS}_t$, restart count, and collapse count $\#\mathrm{collapse} := \#\{t : \mathrm{ESS}_t < 0.1\,N_R\}$).

Table 3. RMSE and degeneracy diagnostics versus ι ($n = 10$, $N_R = 256$, $T = 1.5$, mean ± std over 100 runs).

ι	RMSE	$\overline{\mathrm{ESS}}$	$\min_t \mathrm{ESS}_t$	#restart	#collapse
1	0.316 ± 0.096	170.1	17.4	29.0	0.01
2	0.331 ± 0.096	163.1	21.5	14.0	0.01
3	0.338 ± 0.096	156.3	23.3	9.0	0.01
5	0.349 ± 0.097	146.8	16.8	5.0	0.05
8	0.355 ± 0.100	138.2	9.6	3.0	0.12
10	0.359 ± 0.100	135.1	16.6	2.0	0.10
12	0.358 ± 0.101	133.6	17.6	2.0	0.05
15	0.365 ± 0.102	129.7	14.8	1.0	0.09
20	0.365 ± 0.103	128.1	12.4	1.0	0.10

Empirical guidance for implementations: a Pareto speed–accuracy trade-off

From $\iota = 1$ to $\iota = 10$, the RMSE rises 13.6% and $\overline{\mathrm{ESS}}$ drops 20.5%. We frame the choice of ι as a *Pareto speed–accuracy trade-off*: $\iota = 1$ minimizes RMSE, while larger ι cuts restart operations (from 29 to 1). With $N_R = 256$, the filter stays largely non-degenerate ($\overline{\mathrm{ESS}}$ between 50.1% and 66.4% of N_R, #collapse ≤ 0.12 per run) and frequent restarts ($\iota \leq 5$) give the highest ESS and the lowest RMSE, so a short pilot sweep (Table 3) is recommended per configuration. No universal rule for $\iota(n, N_R)$ is claimed; the optimum depends on the effective dimension and degree of nonlinearity.

5.4. QMC convergence for the kernel integral

The improved Yau–Yau algorithm's prediction step computes

$$\bar{w}(y_i) = \frac{1}{N} \sum_{j=1}^{N} K_{\Delta t}(x_j, y_i)\, w_j,$$

where $\{x_j\}$ are sample points and $K_{\Delta t}$ is the first-order kernel (3.4), a heuristic local Gaussian approximation (Remark 3.1). The deterministic Koksma–Hlawka bound is $\|I_N - I\|_\infty \leq V_{\mathrm{HK}}(K) D_N^*$ with $D_N^* = O((\log N)^{n-1}/N)$. Below we use *randomized* QMC (Owen-scrambled Sobol) and report the average-case RMSE, so we verify a near-N^{-1} trend rather than the KH worst-case bound. For random sampling, the RMSE is $O(N^{-1/2})$.

To directly measure this RQMC average-case error, we isolate the QMC integration error from filtering recursion errors. We fix $r = 2$ ($n = 4$), $\Delta t = 0.1$, Heston parameters $\mu = 0$, $\kappa = 2$, $\theta = 0.25$, $\eta = 0.3$, $R_{12} = 0.5$, $\rho = 0$, and use an adaptive domain with $p_{\mathrm{half}} = 3\sqrt{v_{\min}\Delta t}$ so that the kernel support radius $\sqrt{v_{\min}\Delta t}$ is comparable to the domain width. We fix $M = 200$ evaluation points y_i and compute:

- Reference integral $I_{\mathrm{ref}}(y_i)$ with $N_{\mathrm{ref}} = 2^{16}$ Sobol points and $J = 16$ independent Owen scrambles (rel. reference uncertainty $\approx 4.18 \times 10^{-5}$).
- Approximation $I_N(y_i)$ for $N \in \{64, 128, 256, 512, 1024, 2048\}$ using both Sobol (RQMC) and random (MC) sampling;
- Error $= \mathrm{RMS}_i |I_N(y_i) - I_{\mathrm{ref}}(y_i)|$, averaged over 200 independent trials (each trial uses an independent Owen digital-scramble for Sobol, or an independent `numpy.random.Generator` stream for MC).

Sobol randomization. The 200 Sobol trials use Owen's digital scrambling of the Sobol net, implemented via `scipy.stats.qmc.Sobol` with `scramble=True` and a per-trial seed, so each trial is independent in the usual Monte-Carlo sense while retaining the QMC within-trial structure. The 200 random-sampling trials use independent `numpy.random.Generator(seed=trial_id)` streams.

Table 4 and Figure 1 report the results. A log–log regression gives fitted slopes −0.957 (Sobol) and −0.499 (random), consistent with near-N^{-1} RQMC and near-$N^{-1/2}$ MC trends. At $N = 2048$, the Sobol error (0.089) is about 8.8× smaller than the random error (0.778). Sobol is smaller at every tested N, but only for the tested integrand, not as a general QMC advantage.

Table 4. RQMC vs. MC empirical RMSE convergence for the kernel integral ($n = 4$, $\Delta t = 0.1$, 200 trials).

N	Sobol error	Random error	Ratio (Random/Sobol)
64	2.484	4.423	1.78
128	1.321	3.107	2.35
256	0.880	2.199	2.50
512	0.466	1.570	3.37
1024	0.184	1.111	6.04
2048	0.089	0.778	8.78

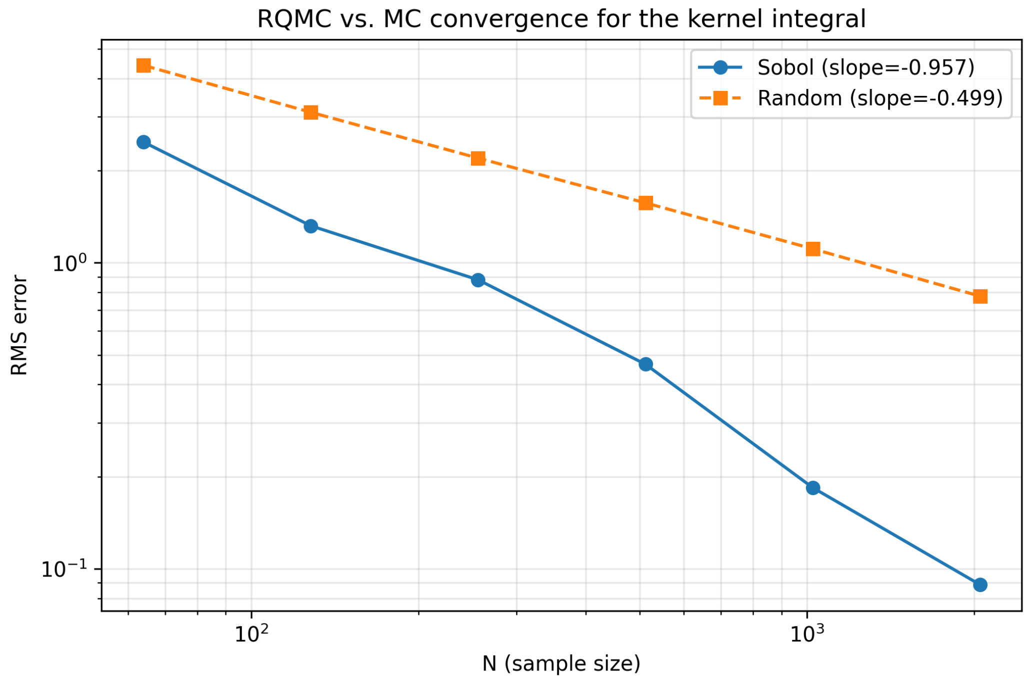

Figure 1. RQMC (Owen-scrambled Sobol) vs. MC (random) empirical RMSE convergence for the kernel integral.

5.5. *Correlation-structure sensitivity: Formula and filter*

A key feature of our model is the generic positive-definite correlation matrix R governing cross-asset price correlations (2.5), extending the standard multivariate Heston assumption $R = I$. To separate the two channels through which R acts, we split the experiment into two parts:

- *5.5.1 Formula sensitivity (oracle R, no filter).* The normalized fixed-target variation $C_A(y_0)$ is computed at three oracle values of R with $(v_{\min}, \eta)$ held fixed. No filter is run, isolating the V_{HK}/C_A dependence on R.
- *5.5.2 Filter sensitivity (filter with R_{filter}, evaluation with R_{true}).* The filter runs with three R_{filter} specifications, but the metric uses the *same* R_{true} throughout, isolating state-estimation error from formula bias.

5.5.1. Formula sensitivity (oracle R)

We fix $r = 2$, $n = 4$, $\eta = 0.3$, $v_{\min} = 0.2$, $v_{\max} = 0.22$, $L_p = 0.02$ (i.e., $\Omega = [-0.02, 0.02]^2 \times [0.2, 0.22]^2$), and use $\Delta t = 0.01$ for the V_{HK} quadrature (the smaller step is required for numerical stability of the high-order derivative integration; this is the same setting used in Section 5.2). We recompute $V_{\mathrm{HK}}(\widetilde{K}_{\Delta t})$ and the resulting C_A at three oracle values of R:

$$R^{(0.9)} = \begin{pmatrix} 1 & 0.9 \\ 0.9 & 1 \end{pmatrix}, \qquad R^{(0.5)} = \begin{pmatrix} 1 & 0.5 \\ 0.5 & 1 \end{pmatrix}, \qquad R^{(0)} = I_2.$$

The likelihood exponent is independent of R; R enters only through $a(x) = \Sigma(x)Q\Sigma(x)^\top$, modifying $K_{\Delta t}$ via a^{-1}. With the upper-face anchor, the values are $C_A^{(0)} = 553.17$ ($R = I_2$), $C_A^{(0.5)} = 1030.61$, $C_A^{(0.9)} = 9.04 \times 10^4$. C_A grows sharply as $R \to \mathbf{1}\mathbf{1}^\top$ because $\det Q \to 0$; thus C_A is not uniformly bounded in R.

5.5.2. Filter sensitivity (filter with R_{filter}, evaluation with R_{true})

We use $r = 2$ assets ($n = 4$) with true correlation $R_{12}^{\text{true}} = 0.9$ (strongly correlated prices), $\Delta t = 0.1$, $T = 2$ (20 steps), and an adaptive domain $p_{\text{half}} = 4\sqrt{v_{\min}\Delta t}$ so that the kernel matrix is non-degenerate. The filter is run with three R_{filter} specifications:

- **Correct** ($R_{\text{filter},12} = 0.9$): matching the true model;
- **Partial** ($R_{\text{filter},12} = 0.5$): partially misspecified;
- **Independent** ($R_{\text{filter}} = I$): fully misspecified.

All three configurations use $N_R = 512$, $\iota = 8$, 1000 runs, and are evaluated against $R_{12}^{\text{true}} = 0.9$. We report (i) price RMSE; (ii) the instantaneous spread-variance error $\sigma_{12,t}^{\text{inst},2} = v_{1,t} + v_{2,t} - 2R_{12}^{\text{true}}\sqrt{v_{1,t}v_{2,t}}$; and (iii) the posterior spread-sd error $\sigma_{12,t}^{\text{post}}$ and its 95% interval coverage.

Table 5 reports the results: Price RMSE and $\sigma_{12,t}^{\text{inst}}$ are nearly invariant to R_{filter} misspecification, whereas $\sigma_{12,t}^{\text{post}}$ is uniformly larger; posterior 95% coverage (83.1%–87.1%) reflects finite-particle bias rather than R misspecification.

Thus R entering the *kernel* changes C_A by $\approx 163\times$, whereas R entering only the *filter* changes metrics by $< 2\%$: The two mechanisms are distinct, and R_{12} must be correctly specified in both.

Table 5. Correlation-structure sensitivity, split design ($r = 2$, $n = 4$, $R_{12}^{\text{true}} = 0.9$, mean ± std over 1000 runs).

Experiment	R	C_A	Price RMSE	$\sigma_{12,t}^{\text{inst}}$ RMSE	$\sigma_{12,t}^{\text{post}}$ RMSE
A: Formula sensitivity (oracle R, no filter)					
$R = I_2$	0.0	553.17	–	–	–
$R^{(0.5)}$	0.5	1030.61	–	–	–
$R^{(0.9)}$	0.9	9.04×10^4	–	–	–
B: Filter sensitivity (filter with R_{filter}, eval with $R_{\text{true}} = 0.9$)					
Correct ($R_{\text{filter},12} = 0.9$)	0.9	–	0.332 ± 0.174	0.020 ± 0.006	0.166 ± 0.059
Partial ($R_{\text{filter},12} = 0.5$)	0.5	–	0.335 ± 0.175	0.020 ± 0.006	0.166 ± 0.054
Independent ($R_{\text{filter}} = I$)	0.0	–	0.338 ± 0.175	0.020 ± 0.006	0.168 ± 0.052

5.6. Spread variance and pairs trading

This section presents an *illustrative case study* designed to demonstrate practical feasibility, not a comprehensive empirical evaluation.

The theoretical analysis in Sections 2–4 uses the linear observation model

$$\mathrm{d}Y_t = HX_t\,\mathrm{d}t + \Gamma_o\,\mathrm{d}W_t^o, \qquad H = [I_r,\, 0_{r\times r}], \tag{5.3}$$

under which Theorems 4.1–4.4 are proved. The real-data study uses only the multivariate return likelihood, not this linear-observation model. Formally, the empirical observation model is

$$\begin{aligned} \Delta Y_t^{\mathrm{ret}} &= \mu_P + \mathrm{diag}(\sqrt{V_t})\,\Delta Z_t^P + \sigma_{\mathrm{ret}}\,\xi_t^{\mathrm{ret}}, \qquad \xi_t^{\mathrm{ret}} \sim \mathcal{N}(0, I_r),\\ \Delta Z_t^P &\sim \mathcal{N}(0, R), \end{aligned} \tag{5.4}$$

where $\Delta Y_t^{\mathrm{ret}}$ is a return observation whose likelihood involves the latent variance V_t but *not* the price coordinates, so the posterior spread variance $\hat{\sigma}_{12,t}^{\mathrm{post}}$ below is a filtering quantity rather than a price-level interval. The theoretical results provide a baseline error bound for the prediction step, while the return update is a model-specific stabilizer.

We apply the filter to daily back-adjusted West Texas Intermediate (WTI) (`CL=F`) and Brent (`BZ=F`) crude oil futures (Yahoo Finance, January 2024–June 2026; 627 trading days, split 439/94/94 training/validation/test). The state dimension is $n = 4$, and the filter uses $\Delta t = 1$ day, $N_R = 512$, restart interval $\iota = 5$ days, and return-likelihood noise $\sigma_{\mathrm{ret}} = 0.003$.

Cointegration. The Engle–Granger test on $\log P_t^{\mathrm{WTI}} = \alpha + \beta \log P_t^{\mathrm{Brent}} + u_t$ rejects no cointegration at the 1% level (ADF −4.27 vs. critical value −3.90), with hedge ratio $\hat{\beta} = 0.988$ (SE 0.006). The spread $s_t = \log P_t^{\mathrm{WTI}} - \hat{\alpha} - \hat{\beta}\log P_t^{\mathrm{Brent}}$ has a mean-reversion half-life $\tau_{1/2} \approx 5.7$ days; out of sample, the residual ADF statistic is −3.66, rejecting a unit root at the 5% level but not at the 1% level (critical values −3.34 and −3.90).

Heston parameters are estimated by Euler quasi-MLE applied to a 20-day realized-variance proxy, with the Feller constraint $2\kappa\theta \geq \eta^2$ enforced in estimation (satisfied strictly; Table 6) and Newey–West standard errors. The reported SEs reflect QMLE sampling variation only (proxy-construction uncertainty is not propagated, a noted limitation). The drift $\hat{\mu}$ is the sample mean log-return, and the correlation $R_{12} = 0.967$ is held constant in the filter.

Table 6. Estimated Heston parameters for WTI and Brent futures via Euler QMLE (Feller constraint enforced). Standard errors in parentheses.

Parameter	WTI	Brent
κ (mean-reversion)	1.042 (0.076)	1.238 (0.093)
θ (long-run var.)	3.72×10^{-4} (2.7×10^{-5})	3.07×10^{-4} (2.3×10^{-5})
η (vol. of vol.)	1.24×10^{-2} (6.7×10^{-4})	1.40×10^{-2} (8.4×10^{-4})
μ (drift)	-2.37×10^{-4}	-2.52×10^{-4}
$2\kappa\theta$	7.76×10^{-4}	7.60×10^{-4}
η^2	1.53×10^{-4}	1.96×10^{-4}
Feller satisfied	Yes	Yes

The filter produces daily posterior variance particles $v_{1,t}^{(i)}$ and $v_{2,t}^{(i)}$ with weights $w_t^{(i)}$. Using the cointegration hedge ratio $\hat{\beta} = 0.988$, the instantaneous spread standard deviation is computed at the particle level as the posterior expectation

$$\hat{\sigma}_{12,t}^{\text{inst}} = \sqrt{\sum_{i=1}^{N_R} w_t^{(i)} \Big(v_{1,t}^{(i)} + \hat{\beta}^2 v_{2,t}^{(i)} - 2\hat{\beta} R_{12} \sqrt{v_{1,t}^{(i)} v_{2,t}^{(i)}} \Big)}, \tag{5.5}$$

where $R_{12} = 0.967$ is the training-period log-return correlation (held fixed). The expectation is over the joint posterior of (v_1, v_2), not a plug-in of posterior means (the concave $\sqrt{v_1 v_2}$ term would otherwise understate the spread variance). Since R_{12} is held fixed, the time variation in $\hat{\sigma}_{12,t}^{\text{inst}}$ is driven entirely by the filter's variance estimates, directly exercising its regime-identification ability. The filter uses hybrid resampling (systematic resampling with jitter, plus local restart when ESS $< N_R/10$), where a 10-day trailing average reduces residual particle noise.

The posterior spread standard deviation (Theorem 4.4 quantity) $\hat{\sigma}_{12,t}^{\text{post}} = \sqrt{\widehat{\text{Var}}(s_{12,t}|\mathcal{Y}_t)}$ measures filtering uncertainty about the current spread level, distinct from the instantaneous volatility $\hat{\sigma}_{12,t}^{\text{inst}}$ that drives trading.

Formally, the strategy decides a position π_{t+1} at the close of day t using information $\mathcal{Y}_t = \sigma(\Delta Y_1^{\text{price}}, \ldots, \Delta Y_t^{\text{price}}, \Delta Y_1^{\text{ret}}, \ldots, \Delta Y_t^{\text{ret}})$, and earns

$$r_{t+1}^{\text{strat}} = \pi_{t+1} \, (s_{12,t+1}^{\text{obs}} - s_{12,t}^{\text{obs}}), \qquad t = T_{\text{train}}, \ldots, T-1.$$

This convention is strictly causal: The position used for the return over $[t, t+1]$ depends only on information available at time t, and the realized return uses only the observed spread change, not the filtered spread change. Within $\mathcal{Y}_t$, the filter posterior is updated only through the return observations (5.4), while the observed price levels enter the signal directly through the cointegration spread s_t (no filtered price coordinate is used in z_t).

We implement a pairs-trading strategy based on a dynamic threshold rule. The trading z-score is $z_t = (s_t - \theta_s)/\hat{\sigma}_{12,t}$, where s_t is the observed cointegration spread, θ_s its training-period mean, and $\hat{\sigma}_{12,t}$ the filter's instantaneous spread-volatility estimate.

- **Enter long position** (buy WTI, sell Brent) when $z_t < -2.0$ and no position is open.
- **Enter short position** (sell WTI, buy Brent) when $z_t > 2.0$ and no position is open.
- **Exit position** when $|z_t| < 0.3$ or after 15 days, whichever comes first.

Parameters are selected by a held-out validation grid search over $z_{\text{entry}} \in \{1.25, 1.5, 1.75, 2.0\}$, $z_{\text{exit}} \in \{0.3, 0.5, 0.7\}$, max_hold $\in \{10, 15\}$, maximizing validation return under a 15% drawdown cap (the test period is never used). Positions are volatility-scaled by $\min(m_t/\hat{\sigma}_{12,t}, 3)$, where m_t is the median of $\hat{\sigma}_{12}$ over the expanding window up to day t starting from the first out-of-sample day (strictly causal), reducing exposure in the turbulent regime and increasing it in the calm regime. The fixed-threshold benchmark trades the same spread with fixed entry bands $|s_t - \theta_s| > 0.020$, exiting when $|s_t - \theta_s| < 0.005$ or after 20 days, at 1× exposure.

Performance metrics are summarized in Table 7: The variance-adaptive statistics are mean ± std over 200 runs under 3% independent and identically distributed (i.i.d.) Gaussian noise on $\hat{\sigma}_{12,t}$ (a volatility-estimation sensitivity, not a confidence interval). On the test period, the strategy attains a cumulative return of 17.17 ± 0.40% (Sharpe 2.318 ± 0.044, max drawdown 5.70 ± 0.17%), versus 18.54% (Sharpe 1.60, drawdown 10.15%) for the fixed-threshold benchmark and 9.99% (drawdown 30.87%) for buy-and-hold WTI. Though lower in absolute return, the strategy achieves the highest Sharpe ratio and the lowest drawdown and, at matched 1× exposure, outperforms the fixed-threshold benchmark, confirming the benefit of volatility scaling. An independent 200-run check with seed 42 gives 17.16 ± 0.37%, 2.318 ± 0.044, 5.68 ± 0.18%.

Table 7. Backtest performance on the 94-day test period (mean ± std over 200 runs with 3% Gaussian noise on $\hat{\sigma}_{12,t}$, seed 20260719). Drawdowns are gross-notional-normalized.

Strategy	Cum. return (%)	Sharpe ratio	Max drawdown (%)
Variance-adaptive (proposed)	17.17 ± 0.40	2.318 ± 0.044	5.70 ± 0.17
Buy-and-hold WTI	9.99	—	30.87
Fixed-threshold (±0.020)	18.54	1.60	10.15
Variance-adaptive @1x	19.45	2.31	6.45
Fixed-threshold @1x	18.54	1.60	10.15

To capture sampling uncertainty beyond volatility-estimation sensitivity, a stationary block bootstrap ([30]; 200 runs, block size 8, seed 42) on the test-period daily returns gives a 95% interval of $[-0.72\%, 40.20\%]$ for the cumulative return and $[-0.100, 4.132]$ for the Sharpe ratio, reflecting the genuine sampling uncertainty of a 94-day test window. The validation-fold dispersion across the $K = 3$ sub-windows is mean Sharpe 0.87, std 0.78, range [0.00, 1.49].

Table 8 stratifies the strategies *ex-post* into a calm (February 2026) and a turbulent (April 2026) month; both spread strategies post small negative returns in the calm month, and the turbulent month contributes the bulk of returns.

The 95% daily value-at-risk (VaR)/conditional value-at-risk (CVaR) are −0.917%/−2.365% (variance-adaptive), −1.703%/−3.575% (fixed-threshold), and −8.563%/−12.598% (buy-and-hold), giving the variance-adaptive strategy the thinnest left tail. Entries are predominantly long WTI / short Brent over this window, so short-WTI profitability remains to be validated on longer samples.

Table 8. Regime-split backtest performance on calm (Feb 2026) and turbulent (Apr 2026) test months (single-run, unperturbed $\hat{\sigma}_{12,t}$).

Sub-period	Strategy	Cum. return (%)	Sharpe	Max drawdown (%)
Calm (Feb 2026)	Variance-adaptive	−0.27	−2.09	0.51
	Fixed-threshold	−0.31	−1.96	0.64
	Buy-and-hold WTI	6.36	—	1.76
Turbulent (Apr 2026)	Variance-adaptive	16.08	5.67	2.49
	Fixed-threshold	10.03	2.14	10.25
	Buy-and-hold WTI	3.58	—	26.89

Transaction costs and risk budget. Exposures are capped at 3× (variance-adaptive) versus

1× (fixed threshold). Returns are quoted per unit of spread notional (1 dollar WTI versus $\hat{\beta}$ dollars Brent; gross notional $1 + \hat{\beta} \approx 1.99$), so drawdowns on this gross-notional basis are roughly half the reported values. Table 9 reports net returns under the one-way cost model $r_{\text{net},t} = r_{\text{gross},t} - \text{cost}_{\text{bps}} \times 10^{-4} \times |e_t - e_{t-1}|$ (with e_t the step-t position) at 0–5 bps; both strategies remain profitable at 5 bps. Both strategies use the same accounting: The equity path is the additive log-spread P&L $E_t = 1 + \sum_{s \le t} e_s \Delta s_s$, and no position is force-liquidated at test end, so turnover counts only within-period changes.

Table 9. Transaction-cost sensitivity at 0–5 bps one-way cost per unit of spread-notional turnover (test period).

Strategy	Cost (bps)	Cum. net (%)	Sharpe	Max DD (%)	Turnover
VA	0.0	17.13	2.305	5.71	8.70
VA	0.5	17.08	2.299	5.72	8.70
VA	1.0	17.04	2.294	5.73	8.70
VA	2.0	16.95	2.282	5.75	8.70
VA	5.0	16.69	2.247	5.81	8.70
FT	0.0	18.54	1.600	10.15	12.0
FT	0.5	18.48	1.595	10.16	12.0
FT	1.0	18.42	1.589	10.17	12.0
FT	2.0	18.30	1.579	10.19	12.0
FT	5.0	17.94	1.548	10.25	12.0

Figure 2 shows $\hat{\sigma}_{12,t}$ (left) and $\hat{v}_{1,t}, \hat{v}_{2,t}$ (right). Table 10 reports $\hat{\sigma}_{12,t}$ in the calm and turbulent subperiods. Since R_{12} is constant, the regime contrast is carried entirely by the filter: $\hat{\sigma}_{12,t}$ rises 1.34× in turbulence, while the posterior variance tracks the market's ≈ 12.8× increase in WTI log-return variance (February 3.2×10^{-4}, April 4.1×10^{-3}) in dampened form.

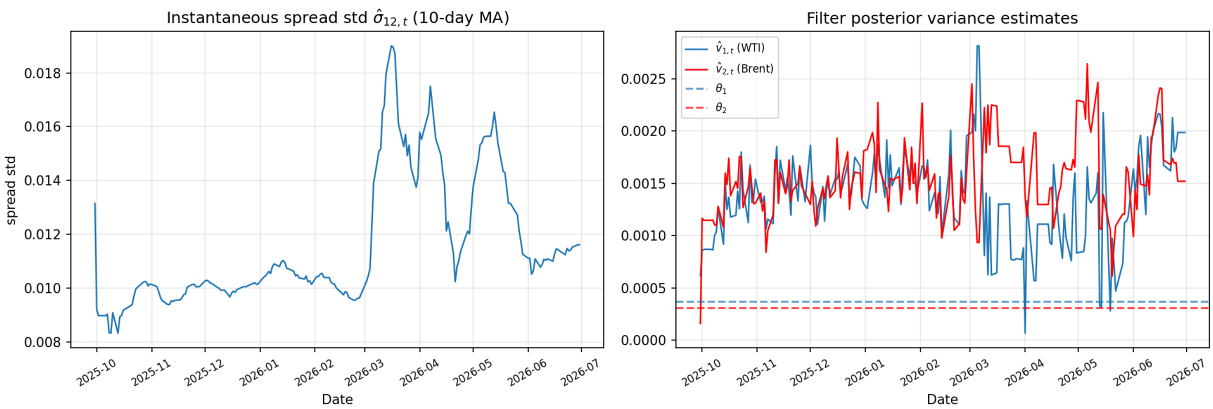

Figure 2. Out-of-sample filter outputs for the WTI/Brent spread (Sep 2025–Jun 2026). Left: Instantaneous spread standard deviation $\hat{\sigma}_{12,t}$ (10-day MA). Right: Posterior variance estimates $\hat{v}_{1,t}$ (WTI) and $\hat{v}_{2,t}$ (Brent), with long-run means θ_1, θ_2 (dashed).

Table 10. Instantaneous spread standard deviation $\hat{\sigma}_{12,t}$ (10-day MA smoothed) in the calm and turbulent test sub-periods. Ratios are relative to the calm baseline (Feb 2026).

Period	median($\hat{\sigma}_{12,t}$)	max($\hat{\sigma}_{12,t}$)	Ratio to calm
Feb 2026 (calm)	0.00923	0.00992	1.00
Apr 2026 (turbulent)	0.01236	0.01274	1.34

During the April 2026 turbulence, the elevated $\hat{\sigma}_{12,t}$ automatically widened the 2.0σ entry bands and the volatility cap reduced position size, while tighter bands during February 2026 permitted larger mean-reversion exposures. The regime identification is attributable solely to the Yau–Yau filter, demonstrating practical feasibility of the variance-driven design.

5.7. Supplementary sensitivity analyses (S5–S7)

This section reports three supplementary sensitivity analyses that complement the main numerical experiments: floating-point precision (S5), anchor convention (S6), and domain and quadrature order (S7).

S5: Floating-point precision sensitivity. Table 11 compares C_A under `float32` vs. `float64` for three parameter settings. The relative difference is below 7×10^{-7} in all cases, confirming that `float64` is more than adequate. The `float32` pipeline rebuilds the kernel, AD, and quadrature in IEEE-754 `float32`; the kernel exponent stays far from the `float32` overflow threshold, so AD differentiates the true kernel.

Table 11. S5: Floating-point precision sensitivity of C_A for three representative parameter settings.

$v_{\min}$	η	$C_A^{\texttt{float64}}$	$C_A^{\texttt{float32}}$	rel. diff.
0.15	0.3	705.2667	705.2667	7.42×10^{-8}
0.20	0.3	553.1705	553.1702	5.31×10^{-7}
0.20	0.2	891.1467	891.1463	4.44×10^{-7}

S6: Anchor sensitivity. Table 12 compares C_A at the upper-face anchor $\boldsymbol{a} = \mathbf{1}$ against the distribution over 60 uniform random interior anchors (seeds 101–160). On average, the random interior anchors yield 21%–24% *larger* C_A than the upper-face anchor, so the upper-face values of Tables 1 and 2 lie near the lower end of that distribution; we adopt $\boldsymbol{a} = \mathbf{1}$ as a deterministic convention without claiming it uniformly minimizes C_A. The qualitative scaling is unaffected by the anchor choice.

S7: Domain half-width and quadrature-order sensitivity. Table 13 reports C_A under four configurations. The quadrature-order variations yield $< 0.3\%$ relative difference, confirming convergence. Doubling the domain half-widths (R_p, R_v) yields an approximately 2.8× larger C_A, so the domain must match the kernel support.

Table 12. S6: Anchor sensitivity of C_A ($\eta = 0.3$, 60 random seeds 101–160): Fixed anchor $a = 1$ versus the random-interior-anchor distribution.

$v_{\min}$	$C_A^{\text{fixed } 1}$	$C_A^{\text{random, mean}}$	$C_A^{\text{random, median}}$	$C_A^{\text{random, } [q_{5\%}, q_{95\%}]}$	$C_A^{\text{random, max}}$	rel. diff. (mean)
0.10	942.63	1164.62	1177.58	$[974.71, 1325.37]$	1396.65	+23.6%
0.20	553.17	667.28	669.07	$[615.47, 710.99]$	727.74	+20.6%
0.30	383.44	467.81	471.43	$[436.46, 491.18]$	496.07	+22.0%

Table 13. S7: Domain half-width and quadrature-order sensitivity of C_A at $v_{\min} = 0.2$, $\eta = 0.3$.

Config	R_p	R_v	Quad. orders	C_A	rel. diff. vs. default
default	0.02	0.01	20/20/10/5	553.17	0.00%
high_quad	0.02	0.01	30/30/15/10	553.91	0.13%
low_quad	0.02	0.01	10/10/5/3	551.89	−0.23%
wide_domain	0.04	0.02	20/20/10/5	1553.89	180.91%

6. Conclusions

We have developed an explicit model-dependent error analysis for the kernel component of the filtering scheme and a truncated-domain spread-moment consequence obtained by composing the $L^\infty \to L^\infty$ operator-norm error bound of [5] (Theorem 5.12) with a normalization lemma ($L^\infty \to L^1$) and a posterior-moment stability lemma ($L^1 \to$ moment), neither of which requires a bounded-Lipschitz density bound. Our primary theoretical contribution is an explicit upper bound for the Hardy–Krause variation of the kernel and likelihood function, derived via a completed-square decomposition and a quadratic analysis of the likelihood exponent. A secondary, model-specific corollary of this composition yields an explicit error bound for truncated posterior spread variance estimators. We have extended the model to accommodate general cross-asset correlation structures, and provided numerical evidence—including QMC convergence verification and correlation-structure sensitivity tests—that support the qualitative polynomial dependence predicted by the theory and demonstrate the importance of correctly specifying R for spread variance estimation. The theoretical bounds should be viewed as conservative qualitative guarantees. The numerical experiments support the predicted polynomial dependence ($\widehat{\beta}_\eta \approx 1.0, \widehat{\beta}_v \approx 0.86$, both well within the bounds of 24 and 12).

The present theorem applies to the idealized continuous-density QMC scheme on the truncated domain Ω, not to the discrete particle-resampling implementation of the empirical study. The theorem also does not establish polynomial complexity in the state dimension, and the real-data return-only observation model is not covered by the linear observation analysis.

An illustrative case study on crude oil futures demonstrates the practical feasibility of the approach. Limitations of the empirical case study include: (a) transaction costs, slippage, and margin requirements are not included in the headline return numbers (a cost-sensitivity table at 0.5–5 bps is reported); (b) the 94-day test window yields only 5 trades for the variance-adaptive strategy (6 for the fixed-threshold benchmark), limiting the statistical stability of VaR, CVaR, and Sharpe estimates;

(c) the variance-adaptive and fixed-threshold strategies are not compared under identical risk budgets (max exposure 3× vs. 1×) and (d) the ex-post regime stratification is descriptive, not predictive.

Future work includes: (i) a comprehensive empirical evaluation across multiple asset classes and extended time periods; (ii) extension to jump-diffusion models to capture sudden price discontinuities; (iii) incorporation of transaction costs and market impact into the filtering framework for direct trading strategy optimization; (iv) investigation of adaptive restart strategies that automatically tune the resampling interval based on real-time diagnostics of the effective sample size; and (v) extension of the explicit QMC error analysis to the empirical observation model (5.4) that includes the return-likelihood component coupled to the latent variance.

Use of Generative-AI tools declaration

The author used DeepSeek-V4-Pro and GLM-5.3 solely for language polishing of the manuscript and supplementary materials (including grammar, spelling, and readability checks). These tools were not used for research design, data analysis, content generation, or reference management. All core research work and conclusions were independently produced by the author, who takes full responsibility for the entire content.

Acknowledgments

This work was supported by the Doctoral Research Start-up Fund of Lishui University (Grant No. QDZK112023048).

Conflict of interest

The author declares no conflicts of interest in this paper.

References

1. R. E. Kalman, R. S. Bucy, New results in linear filtering and prediction theory, *Trans. ASME Ser. D, J. Basic Eng.*, **83** (1961), 95–108. https://doi.org/10.1115/1.3658902
2. A. Doucet, N. de Freitas, N. Gordon, *Sequential Monte Carlo methods in practice*, Springer, New York, 2001. https://doi.org/10.1007/978-1-4757-3437-9
3. N. Chopin, O. Papaspiliopoulos, *An introduction to sequential Monte Carlo*, Springer, Cham, 2020. https://doi.org/10.1007/978-3-030-47845-2
4. S. T. Yau, S. S. T. Yau, Real time solution of the nonlinear filtering problem without memory II, *SIAM J. Control Optim.*, **47** (2008), 163–195. https://doi.org/10.1137/050648353
5. S. T. Yau, Y. S. Niu, An improved Yau-Yau algorithm for high dimensional nonlinear filtering problems, *Pure Appl. Math. Q.*, **21** (2025), 2369–2423. https://doi.org/10.4310/pamq.251222233358
6. S. L. Heston, A closed-form solution for options with stochastic volatility with applications to bond and currency options, *Rev. Financ. Stud.*, **6** (1993), 327–343. https://doi.org/10.1093/rfs/6.2.327

7. M. Avellaneda, J. H. Lee, Statistical arbitrage in the US equities market, *Quant. Finance*, **10** (2010), 761–782. https://doi.org/10.1080/14697680903124632

8. T. Bengtsson, P. Bickel, B. Li, *Curse-of-dimensionality revisited: Collapse of the particle filter in very large scale systems*, In: Probability and Statistics: Essays in Honor of David A. Freedman, Vol. 2, Institute of Mathematical Statistics, Beachwood, OH, 2008, 316–334. https://doi.org/10.1214/193940307000000518

9. C. Snyder, T. Bengtsson, P. Bickel, J. Anderson, Obstacles to high-dimensional particle filtering, *Mon. Weather Rev.*, **136** (2008), 4629–4640. https://doi.org/10.1175/2008MWR2529.1

10. P. Bickel, B. Li, T. Bengtsson, *Sharp failure rates for the bootstrap particle filter in high dimensions*, In: Pushing the Limits of Contemporary Statistics: Contributions in Honor of Jayanta K. Ghosh, Vol. 3, Institute of Mathematical Statistics, Beachwood, OH, 2008, 318–329. https://doi.org/10.1214/074921708000000228

11. P. Rebeschini, R. van Handel, Can local particle filters beat the curse of dimensionality? *Ann. Appl. Probab.*, **25** (2015), 2809–2866. https://doi.org/10.1214/14-AAP1061

12. G. Evensen, *Data assimilation: The ensemble Kalman filter*, Springer, Berlin, 2009. https://doi.org/10.1007/978-3-642-03711-5

13. M. Katzfuss, J. R. Stroud, C. K. Wikle, Understanding the ensemble Kalman filter, *Am. Stat.*, **70** (2016), 350–357. https://doi.org/10.1080/00031305.2016.1141709

14. X. Chen, Z. Sun, Y. Tao, S. S. T. Yau, A uniform framework of Yau–Yau algorithm based on deep learning with the capability of overcoming the curse of dimensionality, *IEEE Trans. Automat. Contr.*, **70** (2025), 339–354. https://doi.org/10.1109/TAC.2024.3424628

15. J. Dick, F. Pillichshammer, *Digital nets and sequences: Discrepancy theory and quasi-Monte Carlo integration*, Cambridge Univ. Press, Cambridge, 2010. https://doi.org/10.1017/CBO9780511761188

16. A. B. Owen, *Multidimensional variation for quasi-Monte Carlo*, In: Contemporary Multivariate Analysis and Design of Experiments (in honour of Professor Kai-Tai Fang's 65th birthday), World Scientific, 2005, 49–74. https://doi.org/10.1142/9789812567765_0004

17. M. B. Giles, B. J. Waterhouse, *Multilevel quasi-Monte Carlo path simulation*, In: Advanced Financial Modelling, Radon Ser. Comput. Appl. Math., Vol. 8, Walter de Gruyter, Berlin, 2009, 165–182. https://doi.org/10.1515/9783110213140.165

18. J. P. Fouque, G. Papanicolaou, K. R. Sircar, *Derivatives in financial markets with stochastic volatility*, Cambridge Univ. Press, Cambridge, 2000.

19. J. Da Fonseca, M. Grasselli, C. Tebaldi, A multifactor volatility Heston model, *Quant. Financ.*, **8** (2008), 591–604. https://doi.org/10.1080/14697680701668418

20. C. Gourieroux, J. Jasiak, R. Sufana, The Wishart autoregressive process of multivariate stochastic volatility, *J. Econometrics*, **150** (2009), 167–181. https://doi.org/10.1016/j.jeconom.2008.12.016

21. D. Duffie, D. Filipović, W. Schachermayer, Affine processes and applications in finance, *Ann. Appl. Probab.*, **13** (2003), 984–1053. https://doi.org/10.1214/aoap/1060202833

22. J. C. Cox, J. E. Ingersoll, S. A. Ross, A theory of the term structure of interest rates, *Econometrica*, **53** (1985), 385–407. https://doi.org/10.2307/1911242

23. M. Jeanblanc, M. Yor, M. Chesney, *Mathematical methods for financial markets*, Springer, London, 2009. https://doi.org/10.1007/978-1-84628-737-4

24. T. E. Duncan, *Probability density for diffusion processes with applications to nonlinear filtering theory*, PhD thesis, Stanford University, 1967.

25. M. Zakai, On the optimal filtering of diffusion processes, *Z. Wahrsch. Verw. Gebiete*, **11** (1969), 230–243. https://doi.org/10.1007/BF00536382

26. S. R. S. Varadhan, Diffusion processes in a small time interval, *Comm. Pure Appl. Math.*, **20** (1967), 659–685. https://doi.org/10.1002/cpa.3160200404

27. A. Friedman, *Stochastic differential equations and applications, Vol. 1*, Academic Press, New York, 1975.

28. N. Ikeda, S. Watanabe, *Stochastic differential equations and diffusion processes*, North-Holland, Amsterdam, 1981. https://doi.org/10.1016/S0924-6509(08)X7014-X

29. F. J. Hickernell, A generalized discrepancy and quadrature error bound, *Math. Comp.*, **67** (1998), 299–322. https://doi.org/10.1090/S0025-5718-98-00894-1

30. D. N. Politis, J. P. Romano, The stationary bootstrap, *J. Am. Stat. Assoc.*, **89** (1994), 1303–1313. https://doi.org/10.1080/01621459.1994.10476870