Model Selection for Diffusion Coefficient Estimation in SDE

Yichuan Huang

▶ **To cite this version:**

Yichuan Huang. Model Selection for Diffusion Coefficient Estimation in SDE. 2025. hal-04901917

HAL Id: hal-04901917
https://hal.science/hal-04901917v1

Preprint submitted on 20 Jan 2025

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

Distributed under a Creative Commons Attribution 4.0 International License

MODEL SELECTION FOR DIFFUSION COEFFICIENT ESTIMATION IN SDE

YICHUAN HUANG
LABORATOIRE MODAL'X, UNIVERSITÉ PARIS NANTERRE, NANTERRE, FRANCE
LABORATOIRE MAP5, UNIVERSITÉ PARIS CITÉ, PARIS, FRANCE
YHUANG@PARISNANTERRE.FR

ABSTRACT. This paper introduces a projection-based estimator for the diffusion coefficient in a stochastic differential equation (SDE) using high-frequency observations of a single trajectory. The method extends existing approaches to noncompact estimation domains, allowing the use of projection spaces spanned by noncompactly supported functions such as Hermite and Laguerre bases. The improvement partly relies on a different decomposition of the squared increments of the processes, which define the approximate regression equation. We propose a data-driven model selection procedure and prove that it enables the estimator to automatically balance squared bias and variance. Numerical experiments confirm its effectiveness across various SDE settings.

1. INTRODUCTION

Consider the diffusion process defined by the stochastic differential equation (SDE)

$$dX_t \;=\; b(X_t)\,dt \;+\; \sigma(X_t)\,dW_t, \quad t \geq 0, \quad X_0 \sim \mu, \tag{1}$$

where $(W_t)_{t\geq 0}$ is a standard Wiener process and μ is a probability measure independent of (W_t). The functions b and σ, referred to respectively as the drift and diffusion coefficients, are assumed to be Lipschitz continuous. Given high-frequency observations $\left(X_{k\Delta_n}\right)_{0\leq k\leq n}$ with sampling intervals $\Delta_n \to 0$ as $n \to \infty$, our goal is to construct a nonparametric estimator of the unknown squared diffusion coefficient $\sigma(\cdot)^2$.

In the literature, two primary methods have been proposed for non-parametric estimation of the coefficients of a diffusion process. The first method relies on kernel-based estimators (see, for instance, Florens-Zmirou [12, 11] for estimating σ^2, and Jiang and Knight [19], Bandi and Phillips [2] for both b and σ^2). These works analyze point-wise consistency and also derive the asymptotic distribution of the proposed estimators. In addition, Jacod [18] surveys the results on convergence rates, addressing both point-wise and integrated convergence for kernel-type estimators.

The second method constructs projection estimators on a chosen finite-dimensional space of basis functions, complemented by a data-driven procedure that determines the dimension adaptively. This strategy has been widely explored for diffusion processes. It was introduced by Genon-Catalot et al. [13], expanded by Hoffmann [15, 16] using wavelet bases, and further simplified by Comte et al. [8] with projection-least-square estimators. However, these projection-based estimators typically only reconstruct the restricted function $\sigma^2_A := \sigma^2\,\mathbb{1}_A$ on a compact set $A \subset \mathbb{R}$. To address non-compact domains, Ella-Mintsa [10] extended the least squares estimator to the entire real line by using N independent copies of the diffusion and letting $N \to \infty$ along with the sample size n. This estimator is essentially a compact supported one, but on expanding intervals $[-\log(N), \log(N)]$. The author establishes risk bounds for both $N = 1$ and $N \propto n$, although adaptivity is only pursued in the latter scenario.

A crucial difference between estimating b and σ^2 is whether the observation horizon $T = n\Delta$ remains fixed or increases with n. When $\Delta = 1/n$ so that $T = 1$ is fixed, a consistent estimation of b from a single path is impossible, while σ^2 can still be estimated consistently in that setting. Approaches to estimate b assume multiple independent copies of the diffusion (with T fixed, e.g., Denis et al. [9], Comte and Genon-Catalot [4]) or let $T \to \infty$ (e.g., Comte and Genon-Catalot [7]), in which case an additional ergodicity condition is imposed. The present paper focuses on σ^2 in a noncompact domain and builds on Comte and Genon-Catalot [7], which treated the noncompact case for b but did not address σ^2.

2020 *Mathematics Subject Classification.* 60J60; 62G05; 62M20.
Key words and phrases. Diffusion Processes; Mixing Processes; Model selection; Non parametric estimation.
This research was partially supported by the Laboratoire d'Excellence (Labex) MME-DII (ANR-11-LBX-0023) under the "Investissements d'Avenir" program.

In this paper, a general family of projection spaces $\{S_m : m \in \mathcal{M}_n\}$ is considered, where $\mathcal{M}_n$ is a model collection. The proposed estimator is given by

$$\widehat{\sigma}^2_m \;=\; \arg\min_{t\in S_m} \frac{1}{n}\sum_{k=1}^{n}\left[\frac{\big(X_{(k+1)\Delta}-X_{k\Delta}\big)^2}{\Delta} - t\big(X_{k\Delta}\big)\right]^2.$$

The truncation of $\widehat{\sigma}^2_m$ is introduced for stability, following a strategy similar to Comte and Genon-Catalot [7]. This truncated projection estimator is shown to be consistent on $\mathbb{R}$ without requiring a compact domain A. In particular, it achieves the nonparametric rate $n^{-s/(s+1)}$ if σ^2 resides in a Sobolev space $W^s(A)$. This is in line with established results in heteroscedastic regression (Comte and Genon-Catalot [6]). Moreover, an automatic, data-driven selection of m is proposed via a suitable penalty function that optimally balances bias and variance. In summary, the three principal novelties are as follows:

- A construction of estimators that employ non-compactly supported bases on $\mathbb{R}$, such as Hermite or Laguerre polynomials, broadening applicability to wide intervals.
- A mechanism that accommodates a broad class of bases (rather than just several fixed bases), imposing only light conditions typically satisfied in many statistical settings.
- An adaptive procedure in the single-trajectory regime ($N = 1$) without restricting itself to a compact domain, improving on Ella-Mintsa [10], which demands $N \propto n$ to achieve adaptivity.

Organization of the paper. Section 2 details the assumptions that ensure the existence and uniqueness of the stationary, ergodic, β-mixing solution to (1). Section 3 presents both nonadaptive and adaptive risk bounds with a discussion of the resulting convergence rates. Section 4 provides numerical experiments illustrating the theoretical claims. Finally, Section 5 contains the proofs of the main results.

Notations. The SDE (1) is assumed to admit a stationary solution $(X_t)_{t\geq 0}$ with invariant density π. The π-norm on $L^2(\mathbb{R}, \pi(x)\,dx)$ is denoted $\|\cdot\|_\pi$. Given observations $\{X_{k\Delta} : 1 \leq k \leq n\}$, the empirical norm is

$$\|t\|_n^2 \;=\; \frac{1}{n}\sum_{k=1}^{n} t\big(X_{k\Delta}\big)^2,$$

with the associated empirical inner product

$$\langle s,\, t\rangle_n \;:=\; \frac{1}{n}\sum_{k=1}^{n} s\big(X_{k\Delta}\big)\, t\big(X_{k\Delta}\big).$$

When taking the expectation of $\|t\|_n^2$, it reduces to the π-norm $\|t\|_\pi^2$. On any subset $A \subset \mathbb{R}$, $L^2(A, dx)$ is endowed with the usual norm $\|\cdot\|$, and the orthogonal projection onto $S_m \subset L^2(A, dx)$ is denoted by f_m if $f \in L^2(A, dx)$. For $m \in \mathbb{N}^+$, the Euclidean norm on $\mathbb{R}^m$ is $\|\cdot\|_{2,m}$. For a symmetric matrix $M \in \mathbb{M}_{m\times m}$, the operator norm $\|M\|_{\mathrm{op}}$ is defined by

$$\|M\|_{\mathrm{op}} \;=\; \sup_{\substack{x\in\mathbb{R}^m\\ \|x\|_{2,m}=1}} \|Mx\|_{2,m} \;=\; \sup_{\substack{x\in\mathbb{R}^m\\ \|x\|_{2,m}=1}} \big|x^\top M x\big|.$$

Equivalently, $\|M\|_{\mathrm{op}}$ is the largest absolute value among the eigenvalues of M.

2. Framework and assumptions

This section specifies the assumptions about the diffusion model and on the projection spaces used to estimate the diffusion coefficient. It also describes the asymptotic sampling framework. The definitions and properties of the relevant function spaces (Besov, Sobolev-Laguerre, and Sobolev-Hermite) are provided.

2.1. Model assumptions.

To guarantee the existence and uniqueness of the diffusion process solution as well as the stationarity and ergodic properties of the process, we impose the following standard assumptions on the model.

Assumption 1. (i) *The drift function b lies in $C^1(\mathbb{R})$, and there exist constants $c_1, c_2 \geq 0$ such that*

$$\big|b'(x)\big| \;\leq\; c_1\big(1+|x|^{c_2}\big) \quad \textit{for all } x \in \mathbb{R}.$$

(ii) *There is a constant c_3 such that*

$$\big|b(x)\big| \;\leq\; c_3\big(1+|x|\big) \quad \textit{for all } x \in \mathbb{R}.$$

(iii) (Dissipativity) *There exist constants $d \geq 0$, $r > 0$, and $R > 0$ such that for all $x \in \mathbb{R}$ with $|x| \geq R$,*

$$\operatorname{sgn}(x)\, b(x) \;\leq\; -r\,|x|^d.$$

Assumption 2. (i) *The diffusion coefficient σ is in $C^2(\mathbb{R})$, and there exist constants $c_4, c_5 \geq 0$ such that*

$$\big|\sigma'(x)\big| \,+\, \big|\sigma''(x)\big| \,\leq\, c_4\big(1+|x|^{c_5}\big) \quad \textit{for all } x \in \mathbb{R}.$$

(ii) *There is a constant $L > 0$ such that for all $x, y \in \mathbb{R}$,*

$$\big|\sigma(x) - \sigma(y)\big| \,\leq\, L\,|x-y|^{1/2}.$$

(iii) (Non-degeneracy) *There is a constant $c_6 > 0$ such that*

$$\big|\sigma(x)\big| \,>\, c_6 \,>\, 0 \quad \textit{for all } x \in \mathbb{R}.$$

Under Assumptions 1(i,ii) and 2(i,ii), SDE (1) admits a unique strong solution for any initial distribution μ. Additionally, using 1(iii) (dissipativity) and 2(iii) (non-degeneracy),

$$V(x) \,=\, \int_0^x \exp\Big(-2\int_0^y \frac{b(z)}{\sigma(z)^2}\,dz\Big)\,dy \,\longrightarrow\, \pm\infty \quad \text{as} \quad x \to \pm\infty,$$

and

$$G \,=\, \int_{-\infty}^{+\infty} \frac{1}{\sigma(x)^2}\exp\Big(2\int_0^x \frac{b(z)}{\sigma(z)^2}\,dz\Big)\,dx \,<\, \infty.$$

A standard argument (see Kutoyants [21], Theorem 1.16) shows that there is a unique invariant measure $\mu(dx) = \pi(x)\,dx$ for (1), where

$$\pi(x) \,=\, \frac{1}{G\,\sigma^2(x)}\exp\Big(2\int_0^x \frac{b(z)}{\sigma(z)^2}\,dz\Big).$$

Assumption 3. *The initial condition satisfies $X_0 \sim \mu$.*

Under Assumptions 1, 2, and 3, the process $(X_t)_{t\geq 0}$ is strictly stationary and ergodic. Moreover, by Proposition 1 in Pardoux and Veretennikov [22], there exist constants $K > 0$, $\nu > 0$, and $\theta > 0$ such that

$$\mathbb{E}\big[\exp\big(\nu\,|X_0|\big)\big] \,<\, +\infty \quad \text{and} \quad \beta_X(t) \,\leq\, K\,\exp\big(-\theta\,t\big), \tag{2}$$

where $\beta_X(t)$ is the β-mixing coefficient at lag t. By the first part of (2), X_0 has finite moments of all orders.

From Gloter [14], Proposition A, for $f = b$ or $f = \sigma$ and any $t > 0$, $k > 0$, $h > 0$, there exist $C = C(k) > 0$ and $\gamma > 0$ such that

$$\mathbb{E}\Big[\sup_{s\in[t,\,t+h]}\big|f(X_s) - f(X_t)\big|^k \,\Big|\, \mathcal{F}_t\Big] \,\leq\, C\,h^{k/2}\,\big(1+|X_t|^{\gamma}\big),$$

where $(\mathcal{F}_t)_{t\geq 0}$ is the filtration generated by W. In particular, there is a constant $C' > 0$ for which

$$\mathbb{E}\Big[\sup_{s\in[t,\,t+h]}\big|f(X_s) - f(X_t)\big|^k\Big] \,\leq\, C'\,h^{k/2}. \tag{3}$$

Assumption 4. *The stationary density π is bounded: $\|\pi\|_\infty < \infty$.*

2.2. Approximation spaces.

The primary goal is to estimate $\sigma_A^2 := \sigma^2\,\mathbb{1}_A$ for a domain $A \subset \mathbb{R}$, which may be a compact interval, $\mathbb{R}^+$, or the entire real line. For this, a family of finite-dimensional subspaces $\{S_m\,;\, m \in \mathbb{N}^+\} \subset L^2(A)$ is considered, each S_m spanned by the orthonormal family $(\varphi_0, \ldots, \varphi_{m-1})$. Specifically,

$$S_m \,=\, \operatorname{span}\{\varphi_0, \ldots, \varphi_{m-1}\}.$$

To control the growth of $\{\varphi_j\}$, the following holds.

Assumption 5. *For each $m \in \mathbb{N}^+$, there is a constant $c_\varphi > 0$ such that*

$$L(m) \,\leq\, c_\varphi^2\,m \quad \textit{with} \quad L(m) \,=\, \sup_{x\in A}\sum_{j=0}^{m-1}\varphi_j(x)^2.$$

Examples of valid bases. The following standard bases, each satisfying Assumption 5, will be important in our analysis. In addition, we recall classical function spaces (Besov, Sobolev-Laguerre, and Sobolev-Hermite) and their associated approximation properties.

1. Trigonometric Basis [TB].

On $[0,1]$, define

$$t_0(x) \;=\; \mathbf{1}_{[0,1]}(x), \quad t_{2j-1}(x) \;=\; \sqrt{2}\,\cos\big(2\pi j\,x\big)\,\mathbf{1}_{[0,1]}(x), \quad t_{2j}(x) \;=\; \sqrt{2}\,\sin\big(2\pi j\,x\big)\,\mathbf{1}_{[0,1]}(x),$$

for $j \geq 1$. Since $|t_j| \leq \sqrt{2}$, it follows that $L(m) \leq 2m$, implying Assumption 5 holds with $c_\varphi^2 = 2$. In many references, these are called the (discrete) Fourier basis functions on $[0,1]$.

Besov-space approximation: If $\sigma^2_{[0,1]}$ is in the Besov space $\mathcal{B}^s_{2,\infty}([0,1])$ with $s > 0$, then by Barron et al. [3], Lemma 12,

$$\big\|\sigma^2_{[0,1]} - \sigma^2_{[0,1],m}\big\|^2 \;\leq\; C(s)\,m^{-2s},$$

where $C(s) > 0$ depends on s. Recall that $\mathcal{B}^s_{2,\infty}([0,1])$ is the set of $f \in L^2([0,1])$ with a finite Besov norm

$$\|f\|_{\mathcal{B}^s_{2,\infty}} \;:=\; \sup_{y>0}\big[y^{-\alpha}\,\omega_r(f,y)_2\big] \;<\; \infty, \quad \text{where } r = \lfloor s \rfloor + 1.$$

Here, $\omega_r(f,y)_2$ is the r-th modulus of smoothness of f in L^2. Thus, for $\sigma^2_{[0,1]} \in \mathcal{B}^s_{2,\infty}([0,1])$, the m-dimensional trigonometric basis approximates $\sigma^2_{[0,1]}$ at rate m^{-s} in the L^2-norm.

2. Laguerre Basis [LB].

On $\mathbb{R}^+$, the Laguerre functions are

$$\ell_j(x) \;=\; \sqrt{2}\,L_j(2x)\,\mathrm{e}^{-x}\,\mathbf{1}_{\{x\geq 0\}}, \quad L_j(u) \;=\; \sum_{k=0}^{j}(-1)^k\binom{j}{k}\,\frac{u^k}{k!}.$$

Since $|\ell_j| \leq \sqrt{2}$, Assumption 5 holds with $c_\varphi^2 = 2$. Smoothness on $\mathbb{R}^+$ is often measured using the Sobolev–Laguerre space (see Thangavelu [25], Chapter 1, and Szegő [24], Chapter 6):

$$W^s_\ell(\mathbb{R}^+, R) \;=\; \Big\{f \in L^2(\mathbb{R}^+)\colon \sum_{j=0}^{\infty} j^s\,\big|\langle f, \ell_j\rangle\big|^2 \;\leq\; R\Big\}.$$

If $\sigma^2_{\mathbb{R}^+} \in W^s_\ell(\mathbb{R}^+, R)$, then

$$\big\|\sigma^2_{\mathbb{R}^+} - \sigma^2_{\mathbb{R}^+,m}\big\|^2 = \sum_{j=m}^{\infty}\big|\langle \sigma^2_{\mathbb{R}^+}, \ell_j\rangle\big|^2 \;\leq\; m^{-s}\sum_{j=m}^{\infty} j^s\,\big|\langle \sigma^2_{\mathbb{R}^+}, \ell_j\rangle\big|^2 \;\leq\; R\,m^{-s}. \tag{4}$$

Hence, the Laguerre basis yields an approximation rate $m^{-s/2}$ in $L^2(\mathbb{R}^+)$.

3. Hermite Basis [HB].

On $\mathbb{R}$, define

$$h_j(x) \;=\; \frac{1}{\sqrt{2^j\,j!\,\sqrt{\pi}}}\,H_j(x)\,\mathrm{e}^{-x^2/2}, \quad H_j(x) \;=\; (-1)^j\,\mathrm{e}^{x^2}\,\frac{d^j}{dx^j}\big(\mathrm{e}^{-x^2}\big).$$

From Indritz [17], one has $|h_j| \leq \pi^{-1/4}$, implying Assumption 5 with $c_\varphi^2 = \pi^{-1/2}$. The Sobolev-Hermite space is introduced by

$$W^s_h(\mathbb{R}, R) \;=\; \Big\{f \in L^2(\mathbb{R})\colon \sum_{j=0}^{\infty} j^s\,|\langle f, h_j\rangle|^2 \;\leq\; R\Big\}.$$

See Thangavelu [25], Chapters 4 and 5, and Szegő [24], Chapter 5, for more details. If $\sigma^2 \in W^s_h(\mathbb{R}, R)$, then $\big\|\sigma^2 - \sigma^2_m\big\|^2 \;\leq\; R\,m^{-s}$, where σ^2_m is the orthogonal projection of σ^2 onto $\mathrm{span}\{h_0, \ldots, h_{m-1}\}$.

2.3. **Asymptotic framework.**
Throughout the paper, the sampling interval is denoted $\Delta = \Delta_n$, and obeys the following constraints:

Assumption 6.

$$\Delta_n \;\to\; 0, \quad \Delta_n\,n^{3/5} \;\to\; +\infty, \quad \textit{and} \quad \Delta_n^2 = o\big(1/n\big) \quad \textit{as } n \to \infty.$$

Hence, while $\Delta_n \to 0$, the total observation horizon $T_n = n\,\Delta_n$ grows unbounded, ensuring sufficient ergodicity-based averaging for consistent nonparametric estimation. For simplicity, $\Delta = \Delta_n$ is often written without subscript.

Observation setup. Let $(X_t)_{t\geq 0}$ be a solution of (1) satisfying Assumptions 1–4, and let

$$X_{k\Delta}, \quad k = 0, \dots, n,$$

be $n+1$ observations recorded in time steps of length Δ. The subsequent sections construct a projection-based estimator for the function σ^2, potentially modified by a small truncation for numerical stability, using the data $(X_0, X_\Delta, \dots, X_{n\Delta})$.

3. Estimation of the diffusion coefficient

In this section, we propose new estimators for the diffusion coefficient σ^2 in a fixed model space S_m, introduce a corresponding model selection procedure, and establish risk bounds for these estimators. The approach builds on the projection ideas in Section 4 of Comte et al. [8], with particular attention to non-compact domains and the use of truncation to stabilize empirical matrices.

3.1. Projection estimator on a fixed space S_m.

Fix $k \in \{1, \dots, n\}$ and define

$$U_{k\Delta} \;=\; \frac{\big(X_{(k+1)\Delta} - X_{k\Delta}\big)^2}{\Delta}.$$

From (1), a regression-type relation holds:

$$U_{k\Delta} \;=\; \sigma(X_{k\Delta})^2 \;+\; V_{k\Delta}^{(1)} \;+\; \overline{V}_{k\Delta} \;+\; R_{k\Delta}, \tag{5}$$

where

$$V_{k\Delta}^{(1)} \;:=\; \sigma^2\big(X_{k\Delta}\big)\Big(\tfrac{(W_{(k+1)\Delta} - W_{k\Delta})^2}{\Delta} - 1\Big)$$

is the principal noise term, while $\overline{V}_{k\Delta}$ and $R_{k\Delta}$ (defined later in the proof of Proposition 1) are additional noise-like and residual terms, respectively, both being negligible.

Projection setup. Let $A \subset \mathbb{R}$ be the domain of interest, and let $\{\varphi_j\}_{j\geq 0} \subset L^2(A, dx)$ be an orthogonal basis of functions supported on A. For each $m \in \mathbb{N}^+$, recall that $S_m \;=\; \operatorname{span}\{\varphi_0, \dots, \varphi_{m-1}\}$. A preliminary estimator $\widehat{\sigma}_m^2$ of σ^2 on A is defined by solving

$$\widehat{\sigma}_m^2 \;=\; \arg\min_{t \in S_m} \gamma_n(t), \quad \text{where} \quad \gamma_n(t) \;=\; \frac{1}{n}\sum_{k=1}^{n}\Big(U_{k\Delta} - t\big(X_{k\Delta}\big)\Big)^2. \tag{6}$$

Introduce the matrices

$$\widehat{\Phi}_m \;=\; \Big[\varphi_j\big(X_{i\Delta}\big)\Big]_{1\leq i\leq n,\; 0\leq j\leq m-1} \;\in\; \mathbb{R}^{n\times m}, \quad \widehat{\Psi}_m \;=\; \Big[\langle \varphi_j, \varphi_k\rangle_n\Big]_{0\leq j,k\leq m-1} \;=\; \frac{1}{n}\,\widehat{\Phi}_m^T\,\widehat{\Phi}_m,$$

and

$$\Psi_m \;=\; \mathbb{E}\big[\widehat{\Psi}_m\big] \;=\; \Big[\int_A \varphi_j(x)\,\varphi_k(x)\,\pi(x)\,dx\Big]_{0\leq j,k\leq m-1}.$$

If $U = (U_\Delta, \dots, U_{n\Delta})^T$ and $\widehat{\Psi}_m$ is invertible, then

$$\widehat{\sigma}_m^2 \;=\; \sum_{j=0}^{m-1} \widehat{a}_j^{(m)}\,\varphi_j, \quad \text{with} \quad \widehat{a}^{(m)} \;=\; \big(\widehat{\Phi}_m^T\,\widehat{\Phi}_m\big)^{-1}\,\widehat{\Phi}_m^T\,U \;=\; \frac{1}{n}\,\widehat{\Psi}_m^{-1}\,\widehat{\Phi}_m^T\,U. \tag{7}$$

Truncation for non-compact A. In order to invert the matrix $\widehat{\Psi}_m$ effectively, we need to ensure that its smallest eigenvalue is sufficiently large. When A is compact, this is typically guaranteed by having a density π that is bounded away from zero. By Proposition 7 in Comte et al. [6], we have $\|\Psi_m^{-1}\|_{\mathrm{op}} \leq \frac{1}{\pi_0}$, where $\pi_0 > 0$ is the lower bound of the density, which implies that the smallest eigenvalue of $\widehat{\Psi}_m$ is bounded below by a deterministic constant with high probability, since we can show that $\widehat{\Psi}_m$ and Ψ_m are close in terms of operator norm (see Proposition 6.1 in Comte et al. [7]). In contrast, for noncompact A, we consider a truncated version of $\widehat{\sigma}_m^2$:

$$\widetilde{\sigma}_m^2 \;=\; \widehat{\sigma}_m^2\,\mathbf{1}_{\Lambda_m}, \tag{8}$$

where

$$\Lambda_m := \Big\{ L(m) \left(\|\widehat{\Psi}_m^{-1}\|_{\mathrm{op}} \vee 1 \right) \le \mathfrak{c} \frac{n\Delta}{\log^2(n\Delta)} \Big\}, \quad \mathfrak{c} = \frac{\theta\big(3\log\big(\frac{3}{2}\big) - 1\big)}{C_0}, \tag{9}$$

with $C_0 \ge 82$ a numerical constant and θ from (2). Lemma 5.1 pinpoints $\mathfrak{c}$, chosen slightly larger than in Comte et al. [7] to accommodate higher precision in estimating σ^2. On Λ_m, the eigenvalues $\{\lambda_i\}_{1\le i\le m}$ of $\widehat{\Psi}_m$ satisfy

$$\inf_{1\le i\le m} \lambda_i \ge \frac{L(m)\,\log^2(n\Delta)}{\mathfrak{c}\,n\,\Delta} > 0.$$

3.2. **Risk bounds for the diffusion coefficient estimator.**
The next proposition decomposes the risk of $\widetilde{\sigma}_m^2$ into bias and variance terms, assuming a suitable stability condition on m:

Assumption 7. *The integer $m \ge 1$ satisfies*

$$L(m) \left(\|\Psi_m^{-1}\|_{\mathrm{op}} \vee 1 \right) \le \frac{\mathfrak{c}\,n\,\Delta}{2\,\log^2(n\,\Delta)} \quad \text{and} \quad m \le n\,\Delta, \tag{10}$$

with $\mathfrak{c}$ as in (9).

Assumption 7 ensures stability, which is automatically satisfied when A is bounded away from zero (see Proposition 4.1 in Comte et al. [5]). In noncompact settings, additional precautions are required (see Propositions 3.4 and 3.5 of Comte et al. [5]). However, it follows from Lemma 5.1 that Λ_m holds with high probability.

Proposition 1. *Under Assumptions 1–7, there are positive constants C, C' (independent of n, m) such that*

$$\mathbb{E}\big[\|\widetilde{\sigma}_m^2 - \sigma_A^2\|_n^2\big] \le 7 \inf_{t\in S_m} \|t - \sigma_A^2\|_\pi^2 + 64\,\frac{\mathrm{Tr}\big[\Psi_m^{-1/2}\,\Psi_{m,\sigma^4}\,\Psi_m^{-1/2}\big]}{n} + C\,\Delta^2 + C'\,\frac{1}{n},$$

where

$$\Psi_{m,\sigma^4} := \Big[\mathbb{E}\big[\varphi_j(X_0)\,\varphi_k(X_0)\,\sigma^4(X_0)\big]\Big]_{0\le j,k\le m-1}.$$

When S_m are nested subspaces (i.e. $S_{m_1} \subset S_{m_2}$ for $m_1 \le m_2$), Lemma 2 in Comte et al. [5] implies $\inf_{t\in S_m} \|t - \sigma_A^2\|_\pi^2 \to 0$ as $m \to \infty$. The terms $C\Delta^2 + \frac{C'}{n}$ go to zero under Assumption 6, so the main challenge is controlling $\mathrm{Tr}\big[\Psi_m^{-1/2}\,\Psi_{m,\sigma^4}\,\Psi_m^{-1/2}\big]$, which is the variance term. If σ is bounded, then

$$\mathrm{Tr}\big[\Psi_m^{-1/2}\,\Psi_{m,\sigma^4}\,\Psi_m^{-1/2}\big] \le \|\sigma^4\|_\infty m$$

by Proposition 3.2 in Comte et al. [7]. In particular, if $\sigma(x) \equiv \sigma$ is constant, then $\Psi_m^{-1/2}\,\Psi_{m,\sigma^4}\,\Psi_m^{-1/2} = \sigma^4\,\mathrm{Id}_m$ and the variance term equals $\sigma^4\,m$.

3.2.1. *Rates in the compact case.*
Suppose A is compact, say $A = [0,1]$. Hoffmann [15], providing a reference for scalar diffusions, uses wavelet bases and achieves an L^2-risk rate of $n^{-2s/(2s+1)}$ if σ_A^2 lies in a Besov space $B^s_{2,\infty}([0,1])$. Our approach recovers the same rate with trigonometric bases:

Trigonometric Basis. Take S_m spanned by $\{\sqrt{2}\cos(2\pi jx), \sqrt{2}\sin(2\pi jx)\}$ plus a constant function.
By Proposition 1,

$$\mathbb{E}\big[\|\widetilde{\sigma}_m^2 - \sigma_{[0,1]}^2\|_n^2\big] \le 7 \inf_{t\in S_m} \|\sigma_{[0,1]}^2 - t\|_\pi^2 + 64\,\frac{\mathrm{Tr}\big[\Psi_m^{-1/2}\,\Psi_{m,\sigma^4}\,\Psi_m^{-1/2}\big]}{n} + C\,\Delta^2 + C'\,\frac{1}{n}.$$

Since $[0,1]$ is compact, $\pi(x)$ is bounded below and above: $0 < \pi_0 \le \pi(x) \le \pi_1 < \infty$. Hence

$$\inf_{t\in S_m} \|\sigma_{[0,1]}^2 - t\|_\pi^2 \le \pi_1 \inf_{t\in S_m} \|\sigma_{[0,1]}^2 - t\|^2.$$

If σ is bounded on $[0,1]$, Proposition 3.2(i) in Comte et al. [7] (adapted with $\sigma^2 \mapsto \sigma^4$) shows

$$\mathrm{Tr}\big[\Psi_m^{-1/2}\,\Psi_{m,\sigma^4}\,\Psi_m^{-1/2}\big] \le \|\sigma_{[0,1]}\|_\infty^4\, m.$$

Finally, if $\sigma_{[0,1]}^2 \in B^s_{2,\infty}([0,1])$, Lemma 12 in Barron et al. [3] yields

$$\inf_{t\in S_m} \|\sigma_{[0,1]}^2 - t\|^2 \lesssim m^{-2s}.$$

Hence,

$$\mathbb{E}\big[\|\widetilde{\sigma}^2_m - \sigma^2_{[0,1]}\|^2_n\big] \lesssim m^{-2s} + \frac{m}{n} + \Delta^2 + \frac{1}{n}.$$

By Assumption 6, Δ^2 is negligible compared to $1/n$. Under the compactness of A, Assumption 7 reduces to

$$m \lesssim \frac{n\Delta}{\log^2(n\Delta)}.$$

For functions with regularity $s > 3/4$, then under $n^{3/5}\Delta \to \infty$ (Assumption 6), we can choose $m_{\text{opt}} \approx n^{1/(2s+1)}$, that satisfies Assumption 7 and balances m^{-2s} and m/n, resulting in the rate $n^{-\frac{2s}{2s+1}}$. This matches the result in Hoffmann [15].

3.2.2. *Rates in the non compact case.*
When A is noncompact, for instance $A = \mathbb{R}^+$ or $A = \mathbb{R}$, we estimate σ^2_A using **[LB]** in the first case and **[HB]** in the second. We evaluate the risk of the estimator by adapting the same approach as previously. For example, when $A = \mathbb{R}^+$, we can bound the bias term using Assumption 4:

$$\inf_{t\in S_m} \left\|\sigma^2_{\mathbb{R}^+} - t\right\|^2_\pi \leq \|\pi\|_\infty \inf_{t\in S_m} \left\|\sigma^2_{\mathbb{R}^+} - t\right\|^2. \tag{11}$$

We also assume that σ^2 is bounded (which is trivially true for processes such as the Ornstein-Uhlenbeck process). This allows us to recover the bound for the variance term from the previous section:

$$\operatorname{Tr}\left[\Psi_m^{-1/2}\Psi_{m,\sigma^4}\Psi_m^{-1/2}\right] \leq \|\sigma_{\mathbb{R}^+}\|^4_\infty m. \tag{12}$$

Finally, if we assume that $\sigma^2_{\mathbb{R}^+}$ belongs to a Sobolev-Laguerre space ball $W^s_\ell(\mathbb{R}^+, R)$, we can use (4) to bound the bias term:

$$\inf_{t\in S_m} \left\|\sigma^2_{\mathbb{R}^+} - t\right\|^2 = \|\sigma^2_{\mathbb{R}^+} - \sigma^2_{\mathbb{R}^+,m}\|^2 \leq Cm^{-s}. \tag{13}$$

For functions with regularity $s > 3/2$, then under $n^{3/5}\Delta \to \infty$ (Assumption 6), we can choose $m_{\text{opt}} \approx n^{1/(s+1)}$, that satisfies Assumption 7 and achieves the rate $n^{-s/(s+1)}$.

Comparison to standard heteroscedastic regression.

The regression model

$$Y_k = f(X_k) + g(X_k)\,\varepsilon_k, \quad \mathbb{E}[\varepsilon_k] = 0, \quad \operatorname{Var}[\varepsilon_k] = 1,$$

is studied in Comte et al. [6], where $g(\cdot)$ is a nuisance function that describes the level of noise. On the surface, our model (5) is similar: $U_{k\Delta}$ acts as a response Y_k, $\sigma^2(\cdot)$ parallels $f(\cdot)$, and additional terms $(\overline{V}_{k\Delta}, R_{k\Delta})$ are reminiscent of extra noise.

However, two significant differences arise:

- Process-driven residuals $\overline{V}_{k\Delta}$ and $R_{k\Delta}$ do not simply vanish under expectation but are only of smaller order (often on the scale of Δ). This is typically enough to remain negligible for large n under Assumption 6, yet they complicate the analysis by adding correlated noise-like components.
- The function σ^2 to be estimated is not just the mean structure but also explicitly determines the variance part of the model. That is, $U_{k\Delta}$ depends on $\sigma^2(X_{k\Delta})$ in a multiplicative way, so there is no separation of the "variance function" $g(\cdot)$ from the "mean function" $f(\cdot)$.

In Comte et al. [6], the notion of regularity involves an $L^2(\mathbb{R}^+, \pi(x)\,dx)$-based Sobolev space $\overline{W}^s_\pi(\mathbb{R}^+, R)$, where the "bias" term is controlled via the orthogonal projection in that weighted norm.

$$\overline{W}^s_\pi(\mathbb{R}^+, R) = \left\{h \in \mathbb{L}^2(\mathbb{R}, \pi(x)\,dx) : \forall \ell \geq 1, \|h - h^\pi_\ell\|^2_\pi \leq R\ell^{-s}\right\},$$

with h^π_ℓ the $\mathbb{L}^2(A, \pi(x)\,dx)$ orthogonal projection of h on S_ℓ, and π is the stationary density of X.

By the boundedness of π, if $\sigma^2_{\mathbb{R}^+} \in W^s_\ell(\mathbb{R}^+, R)$ (Laguerre space in the unweighted sense), then

$$\|\sigma^2_{\mathbb{R}^+} - t\|^2_\pi \leq \|\pi\|_\infty \|\sigma^2_{\mathbb{R}^+} - t\|^2,$$

so $\sigma^2_{\mathbb{R}^+}$ also belongs to the $\|\cdot\|_\pi$-Sobolev-type class $\overline{W}^s_\pi(\mathbb{R}^+, \|\pi\|_\infty R)$. Consequently, the $n^{-s/(s+1)}$ rate found here aligns the rates for heteroskedastic regression in Comte et al. [6].

3.3. **Model selection.**

Throughout this section, assume also that:

Assumption 8. *For $m_1, m_2 \in \mathbb{N}^+$ with $m_1 \le m_2$, we have $S_{m_1} \subset S_{m_2}$.*

Constrained Model Collection. Define

$$\mathcal{M}_{n\Delta} := \Big\{ m \in \mathbb{N}^+ : \; c_\varphi^2\, m \left(\|\Psi_m^{-1}\|_{\mathrm{op}}^2 \vee 1 \right) \le \frac{\mathfrak{d}}{4} \frac{n\,\Delta}{\log^2(n\Delta)} \Big\}, \quad \mathfrak{d} = \frac{\theta}{8\, C_0 \big(\|\pi\|_\infty \vee (1 + \frac{1}{3}) \big)}, \tag{14}$$

where θ is from (2), $C_0 \ge 84$ is the same constant as in $\mathfrak{c}$ of (9), and the factor of 4 ensures a stricter constraint than (9). To enable an automatic selection of m, we introduce an *empirical* version of $\mathcal{M}_{n\Delta}$:

$$\widehat{\mathcal{M}}_{n\Delta} := \Big\{ m \in \mathbb{N}^+ : \; c_\varphi^2\, m \left(\|\widehat{\Psi}_m^{-1}\|_{\mathrm{op}}^2 \vee 1 \right) \le \mathfrak{d} \frac{n\,\Delta}{\log^2(n\Delta)} \Big\}. \tag{15}$$

Given that $m \left(\|\Psi_m^{-1}\|_{\mathrm{op}} \vee 1 \right) \le m \left(\|\Psi_m^{-1}\|_{\mathrm{op}}^2 \vee 1 \right)$, it follows that $\widehat{\sigma}^2_m = \widetilde{\sigma}^2_m$. Then the *data-driven* choice of m is

$$\widehat{m} = \arg \min_{m \in \widehat{\mathcal{M}}_{n\Delta}} \Big\{ -\|\widehat{\sigma}^2_m\|_n^2 + \kappa\, c_\varphi^2\, s^2\, \frac{m\, \|\widehat{\Psi}_m^{-1}\|_{\mathrm{op}}}{n} \Big\}, \quad s^4 = \mathbb{E}\big[\sigma^4(X_0)\big], \tag{16}$$

where κ is a numerical constant. Intuitively, $-\|\widehat{\sigma}^2_m\|_n^2$ approximates the squared bias, and the second term controls the variance (cf. the proof of Proposition 3.2(iii) in [7]). Note that $\widehat{m}$ is chosen only among the indices $m \in \widehat{\mathcal{M}}_{n\Delta}$, so that $\widehat{\sigma}^2_{\widehat{m}} = \widetilde{\sigma}^2_{\widehat{m}}$ for those m.

Theorem 3.1. *Under Assumptions 1–8, there is a numerical constant $\kappa_0 > 0$ such that for all $\kappa \ge \kappa_0$,*

$$\mathbb{E}\big[\|\widehat{\sigma}^2_{\widehat{m}} - \sigma^2_A\|_n^2\big] \le C \inf_{m \in \mathcal{M}_{n\Delta}} \Big(\inf_{t \in S_m} \|t - \sigma^2_A\|_\pi^2 + \mathrm{pen}(m) \Big) + C'\, \Delta^2 + C''\, \frac{1}{n},$$

where

$$\mathrm{pen}(m) = \kappa\, c_\varphi^2\, \mathbb{E}\big[\sigma^4(X_0)\big]\, \frac{m\, \|\Psi_m^{-1}\|_{\mathrm{op}}}{n}.$$

Hence, the adaptive estimator $\widehat{\sigma}^2_{\widehat{m}}$ automatically balances bias and variance, up to small remainder terms. A practical implementation requires estimating $\mathbb{E}[\sigma^4(X_0)]$ (see the next section).

4. Simulation study

This section evaluates the performance of the proposed estimation methods in various diffusion processes. In Examples 1, 3, and 4, the processes are mean-reverting to ensure stationarity. Concretely, the drift function in (1) is

$$b(x) = -\lambda\,(x - \mu),$$

where $\lambda > 0$ is the mean reversion rate and μ is the long-term mean. In Example 2, stationarity is inherited from the setting in Example 3. Our primary objective is to estimate σ, or equivalently σ^2.

Example 1: Ornstein–Uhlenbeck (OU) Process.
Parameters: $\lambda = 1$, $\mu = 0$, $\gamma = 2$.

$$dX_t = -\lambda\,(X_t - \mu)\,dt + \sigma(X_t)\,dW_t, \quad \text{where} \quad \sigma(x) = \gamma.$$

Example 2: Nonlinear Diffusion Process.
Parameters: $\lambda = 1$, $\mu = 0$, $\gamma = 2$.

$$dY_t = (1 - Y_t^2) \Big[\lambda\big(\mu - \tanh^{-1}(Y_t)\big) - \gamma^2\, Y_t \Big]\, dt + \sigma(Y_t)\, dW_t, \quad \text{where} \quad \sigma(y) = \gamma\,\big(1 - y^2\big).$$

Example 3: Hyperbolic-Like Diffusion Process.
Parameters: $\lambda = 2$, $\mu = 0$, $\gamma = \frac{1}{\sqrt{2}}$.

$$dU_t = -\lambda\,(U_t - \mu)\,dt + \sigma(U_t)\,dW_t, \quad \text{where} \quad \sigma(u) = \gamma\,\sqrt{1 + u^2}.$$

Example 4: CIR Process.
Parameters: $\lambda = 2$, $\mu = \frac{3}{8}$, $\gamma = 1$, with $\gamma \le 2\,\lambda\,\mu$.

$$dV_t = -\lambda\,(V_t - \mu)\,dt + \sigma(V_t)\,dW_t, \quad \text{where} \quad \sigma(v) = \gamma\,\sqrt{(v)_+}.$$

Examples 1 and 3 adhere to **Assumptions** 1–4, thus allowing the full theoretical framework to apply. **Example 1**, an Ornstein–Uhlenbeck (OU) process, is a classical mean-reverting diffusion widely used in finance, physics, and econometrics. Because it is linear with a closed-form solution, exact simulation is straightforward:

$$X_{k\Delta} \;=\; X_{(k-1)\Delta}\,\mathrm{e}^{-2\Delta} \;+\; \sqrt{2\left(1-\mathrm{e}^{-2\Delta}\right)}\,\varepsilon_k,$$

Where $\{\varepsilon_k\}$ are i.i.d. standard normal random variables. From the sample $\{X_t\}$, we construct a nonlinear diffusion by defining $Y_t \;=\; \tanh(X_t)$. By Itô's lemma, the process $\{Y_t\}$ then satisfies the model in **Example 2**. We use a trigonometric basis supported on $[aX, bX]$ for **Examples 1 and 2**, where aX and bX are the 2% and 98% quantiles of the observations, respectively. This approach is a slight modification of **[TB]** with an adjusted period. Although not all assumptions of **Example 2** are strictly satisfied, the process is stationary and exhibits geometric β-mixing. The same behavior is observed in **Example 4**.

Example 3 is a hyperbolic-like diffusion whose diffusion coefficient grows similarly to a hyperbolic function for large $|x|$. It can be checked that U_t meets Assumptions 1–4. Simulation uses the Euler–Maruyama scheme:

$$U_{(k+1)\Delta} \;=\; U_{k\Delta} \;-\; \Delta\,U_{k\Delta} \;+\; \sqrt{1+U_{k\Delta}^2}\,\sqrt{\Delta}\,\varepsilon_k.$$

Here, $\sqrt{1+\cdot^2}$ is approximated with an $\mathbb{R}$-supported basis [HB]. **Example 4** uses the restricted CIR process so that V_t stays positive, making the Laguerre basis [LB] suitable. For simulation, the Euler–Maruyama method is replaced by the implicit Euler scheme from Alfonsi [1]:

$$V_{(k+1)\Delta} \;=\; \left(\frac{\frac{\gamma}{2}\sqrt{\Delta}\,\varepsilon_k \;+\; \sqrt{V_{k\Delta}} \;+\; \sqrt{\left(\frac{\gamma}{2}\sqrt{\Delta}\,\varepsilon_k \;+\; \sqrt{V_{k\Delta}}\right)^2 \;+\; 4\Big(1+\frac{\lambda\Delta}{2}\Big)\,\frac{\lambda\mu-\frac{\gamma^2}{4}}{2}\,\Delta}}{2\big(1+\frac{\lambda\Delta}{2}\big)}\right)^2,$$

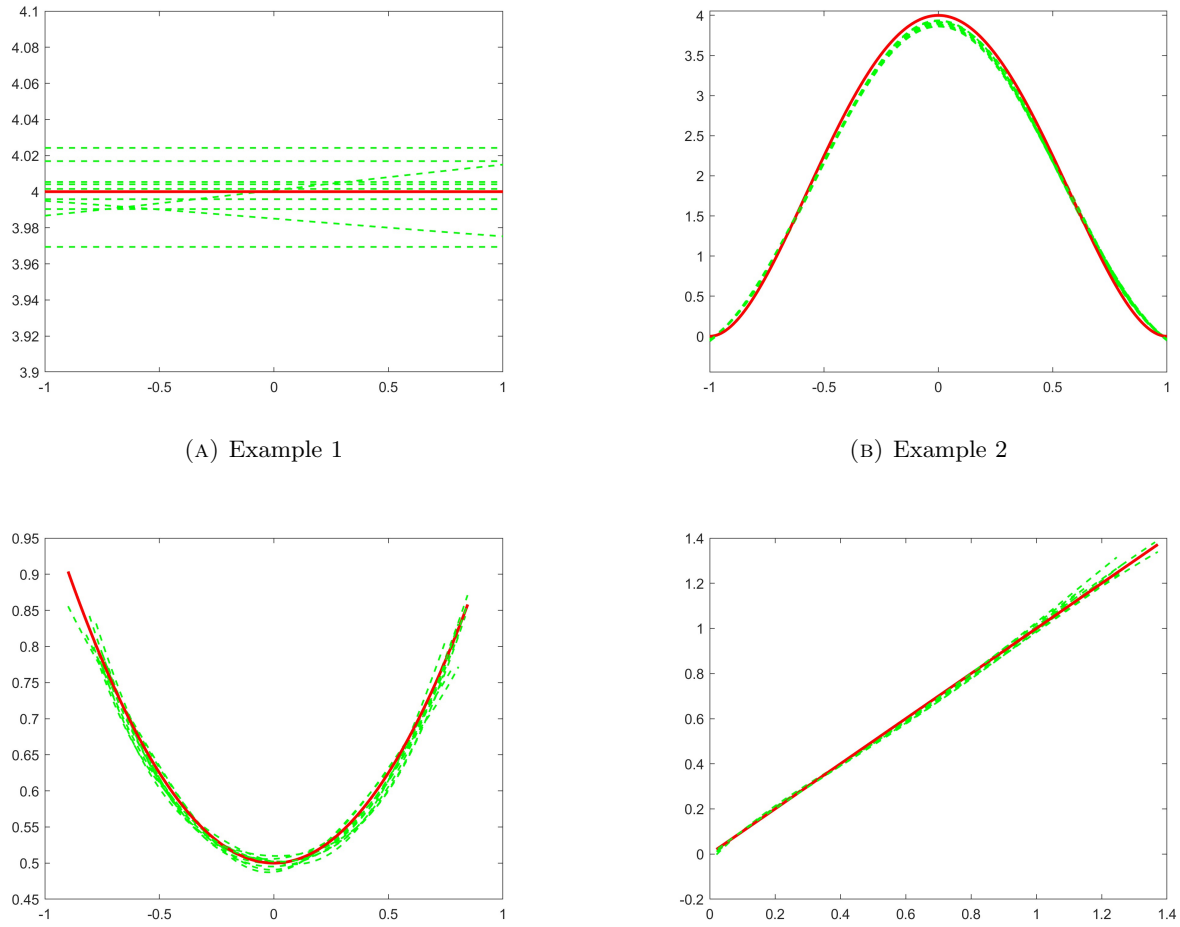
which remains well-defined for large n whenever $\gamma^2 \le 2\,\lambda\,\mu$.

Parameter Tuning and Implementation.
The constant κ for the selection procedure (see (16)) is chosen via standard calibration: $\kappa = 5\times 10^{-2}$ for [TB], and $\kappa = 2\times 10^{-4}$ for [LB] and [HB]. The constant $\mathfrak{d}$ is set to 10^{12}. To handle the quantity $s^4 = \mathbb{E}[\sigma^4(X_0)]$, we replace it with the estimator

$$\widehat{s}^4 \;=\; \frac{1}{2n}\sum_{k=1}^{n}\Big[U_{k\Delta} \;-\; \widehat{\sigma}^2_{\widehat{m}_{\max}}(X_{k\Delta})\Big]^2,$$

where $\widehat{m}_{\max}$ is the maximal dimension in $\widehat{\mathcal{M}}_{n\Delta}$. This choice follows from the relation $\mathbb{E}\big[(V^{(1)}_{k\Delta})^2\big] = 2\,\mathbb{E}\big[\sigma^4(X_0)\big]$, where $V^{(1)}_{k\Delta}$ is defined in (5). In Table 1, we compare $\widehat{s}^4$ to

$$\widetilde{s}^4 \;=\; \frac{1}{n}\sum_{k=1}^{n}\sigma^4\big(X_{k\Delta}\big).$$



(A) Example 1

(B) Example 2

(C) Example 3

(D) Example 4

FIGURE 1. Estimated diffusion coefficient curves for the four examples. Green lines: estimated coefficients; red lines: true coefficients. The bases used are (a) [TB], (b) [TB], (c) [HB], and (d) [LB]. Parameters: $n = 100{,}000$, $\Delta = 0.004$.

	Example 3		Example 4	
	$\widehat{s^4}$	$\widetilde{s^4}$	$\widehat{s^4}$	$\widetilde{s^4}$
$n = 25{,}000,\ \Delta = 0.01$	0.34 (0.15)	0.34 (0.14)	1.66 (0.24)	1.67 (0.19)
$n = 100{,}000,\ \Delta = 0.004$	0.34 (0.01)	0.34 (0.01)	1.57 (0.01)	1.64 (0.01)

TABLE 1. Mean values (and standard deviations in parentheses) of 100 samples for the pseudo-estimators $\widehat{s^4}$ and $\widetilde{s^4}$, with different n and Δ. All entries are rounded to two decimal places.

Numerical Results.

The tables below (Table 2 and Table 3) summarize the estimates of σ^2 under varying n and Δ for Examples 3 and 4. Each table shows the mean integrated squared error (MISE), the selected dimension $\widehat{m}$, and the maximal dimension $\widehat{m}_{\max}$, all averaged over 100 replications (standard deviations in parentheses).

In both cases, the product $n\Delta$ (ranging roughly between 250 and 1000) plays a crucial role. When $n\Delta$ is moderate, the estimation of σ^2 remains stable and accurate, consistent with the fact that, in general, the diffusion coefficients σ^2 can be reliably estimated from the observed data over relatively short time intervals.

For instance, in the first table, $n = 100{,}000$ and $\Delta = 0.004$ yield $n\Delta = 400$, with MISE around 0.02. Moreover, the chosen dimension remains within a reasonable range.

As n grows large, the MISE becomes less sensitive to Δ. Once n is large enough, reducing Δ further makes little difference in the accuracy of the estimation. In Table 2, for $n = 250{,}000$, the MISE is extremely low and remains nearly constant in all tested Δ (roughly 0.0122 to 0.0133). Thus, for large samples, the estimator performance is essentially Δ-invariant, in line with the theoretical risk bounds that are primarily dependent on n.

n	Δ		
	0.04	0.01	0.004
10,000	MISE = 0.22 (0.17)	MISE = 0.34 (0.41)	MISE = 4.15 (2.44)
	$\widehat{m} = 5.33(0.85)$	$\widehat{m} = 3.00(0.00)$	$\widehat{m} = 1.99(0.10)$
	$\widehat{m}_{\max} = 8.45(0.56)$	$\widehat{m}_{\max} = 3.00(0.00)$	$\widehat{m}_{\max} = 2.00(0.00)$
25,000	MISE = 0.12 (0.04)	MISE = 0.06 (0.05)	MISE = 0.29 (0.31)
	$\widehat{m} = 5.81(0.98)$	$\widehat{m} = 5.27(0.66)$	$\widehat{m} = 3.00(0.00)$
	$\widehat{m}_{\max} = 9.51(0.92)$	$\widehat{m}_{\max} = 7.00(0.00)$	$\widehat{m}_{\max} = 3.00(0.00)$
100,000	MISE = 0.08 (0.02)	MISE = 0.02 (0.03)	MISE = 0.02 (0.02)
	$\widehat{m} = 6.65(1.06)$	$\widehat{m} = 6.11(1.00)$	$\widehat{m} = 5.72(0.93)$
	$\widehat{m}_{\max} = 10.66(0.92)$	$\widehat{m}_{\max} = 9.52(0.78)$	$\widehat{m}_{\max} = 8.51(0.56)$
250,000	MISE = 0.01 (0.01)	MISE = 0.01 (0.01)	MISE = 0.01 (0.01)
	$\widehat{m} = 7.23(0.98)$	$\widehat{m} = 6.91(0.93)$	$\widehat{m} = 6.38(1.12)$
	$\widehat{m}_{\max} = 10.77(0.95)$	$\widehat{m}_{\max} = 10.33(1.02)$	$\widehat{m}_{\max} = 9.47(0.90)$

TABLE 2. Results for Example 3 (hyperbolic-like diffusion). MISE ($\times 100$), $\widehat{m}$, and $\widehat{m}_{\max}$, averaged over 100 replications with standard deviations in parentheses. All entries are rounded to two decimal places.

n	Δ		
	0.04	0.01	0.004
10,000	MISE = 0.31 (0.31)	MISE = 0.25 (0.36)	MISE = 3.78 (3.76)
	$\widehat{m} = 4.02(0.28)$	$\widehat{m} = 3.00(0.00)$	$\widehat{m} = 2.00(0.00)$
	$\widehat{m}_{\max} = 5.21(0.41)$	$\widehat{m}_{\max} = 3.00(0.00)$	$\widehat{m}_{\max} = 2.00(0.00)$
25,000	MISE = 0.22 (0.11)	MISE = 0.07 (0.07)	MISE = 0.24 (0.40)
	$\widehat{m} = 4.03(0.17)$	$\widehat{m} = 4.00(0.20)$	$\widehat{m} = 3.00(0.00)$
	$\widehat{m}_{\max} = 5.63(0.49)$	$\widehat{m}_{\max} = 5.16(0.37)$	$\widehat{m}_{\max} = 3.00(0.00)$
100,000	MISE = 0.05 (0.02)	MISE = 0.03 (0.02)	MISE = 0.02 (0.02)
	$\widehat{m} = 4.02(0.19)$	$\widehat{m} = 4.11(0.31)$	$\widehat{m} = 4.11(0.31)$
	$\widehat{m}_{\max} = 5.73(0.25)$	$\widehat{m}_{\max} = 5.36(0.48)$	$\widehat{m}_{\max} = 5.36(0.48)$
250,000	MISE = 0.02 (0.01)	MISE = 0.02 (0.01)	MISE = 0.01 (0.01)
	$\widehat{m} = 4.33(0.13)$	$\widehat{m} = 4.17(0.38)$	$\widehat{m} = 4.17(0.38)$
	$\widehat{m}_{\max} = 6.07(0.40)$	$\widehat{m}_{\max} = 5.96(0.20)$	$\widehat{m}_{\max} = 5.57(0.50)$

TABLE 3. Results for Example 4 (CIR process). MISE, $\widehat{m}$, and $\widehat{m}_{\max}$, averaged over 100 replications (standard deviations in parentheses). All entries are rounded to two decimal places.

5. Proofs

This section presents the proofs of the main results from Section 3. In what follows, let C and c represent constants that do not depend on n, Δ, or m. These constants may change from line to line. The notation $x \lesssim y$ indicates that $x \le c\, y$ for some universal constant c.

5.1. Proof of Proposition 1.

The approach aligns with the proof of Proposition 3.1 in [7]. Define the random event

$$\Omega_m := \Big\{ \big| \tfrac{\|t\|_n^2}{\|t\|_\pi^2} - 1 \big| \le \tfrac{1}{2},\ \forall\, t \in S_m \Big\}. \tag{17}$$

On Ω_m, for functions in S_m, the empirical norm $\|\cdot\|_n$ and the π-norm $\|\cdot\|_\pi$ are equivalent, since

$$\tfrac{1}{2}\,\|t\|_\pi \ \le\ \|t\|_n \ \le\ 2\,\|t\|_\pi \quad \text{for all } \ t \in S_m.$$

The error decomposition is given by

$$\|\widetilde{\sigma}_m^2 - \sigma_A^2\|_n^2 \ \le\ \|\sigma_A^2\|_n^2\, \mathbb{1}_{(\Lambda_m)^c} \ +\ \|\widehat{\sigma}_m^2 - \sigma_A^2\|_n^2\, \mathbb{1}_{\Lambda_m \cap (\Omega_m)^c} \ +\ \|\widehat{\sigma}_m^2 - \sigma_A^2\|_n^2\, \mathbb{1}_{\Lambda_m \cap \Omega_m},$$

since $\widetilde{\sigma}_m^2$ is defined as zero on $(\Lambda_m)^c$. The expectation of each of these three terms is bounded in turn. The following lemma is used repeatedly.

Lemma 5.1. *Under the assumptions of Proposition 1, if m satisfies the stability condition (10), then*

$$\mathbb{P}\big((\Lambda_m)^c\big) \ \le\ \mathbb{P}\big((\Omega_m)^c\big) \ \lesssim\ \frac{1}{(n\Delta)^6}.$$

The proof of Lemma 5.1 follows from Proposition 6.1(i) and Lemma 6.1 in [7]. The difference in order $n\Delta$ is due to the setting $q_n\Delta = 7\log(n\Delta)$ (rather than $6\log(n\Delta)$ as in [7]), reflecting a higher precision in the estimation of the diffusion coefficient. Consequently, a larger constant $\mathfrak{c}$ appears in (10).

Step 1. Bound on $\mathbb{E}\big[\|\sigma_A^2\|_n^2\, \mathbb{1}_{(\Lambda_m)^c}\big]$.

$$\mathbb{E}\big[\|\sigma_A^2\|_n^2\, \mathbb{1}_{(\Lambda_m)^c}\big] \le C\, \mathbb{E}\big[\sigma^8(X_0)\big]^{\frac{1}{2}}\, \mathbb{P}\big((\Lambda_m)^c\big)^{\frac{1}{2}} \ \le\ C'\, \frac{1}{(n\Delta)^3} \ \lesssim\ \frac{1}{n},$$

where Cauchy–Schwarz, Lemma 5.1, and the condition $\Delta\, n^{5/3} \to +\infty$ are applied in turn.

Step 2. Bound on $\mathbb{E}\big[\|\widehat{\sigma}_m^2 - \sigma_A^2\|_n^2\, \mathbb{1}_{\Lambda_m \cap \Omega_m}\big]$.

The proof requires a precise expression for noise terms and negligible residuals in the regression relation (5) used to estimate σ^2. By (1) and Fubini's theorem,

$$U_{k\Delta} = \sigma(X_{k\Delta})^2 + V_{k\Delta} + R_{k\Delta},$$

where

$$V_{k\Delta} = V_{k\Delta}^{(1)} + V_{k\Delta}^{(2)} + V_{k\Delta}^{(3)} + V_{k\Delta}^{(4)},$$

with

$$\begin{aligned}
V_{k\Delta}^{(1)} &:= \sigma^2(X_{k\Delta})\Big(\tfrac{(W_{(k+1)\Delta} - W_{k\Delta})^2}{\Delta} - 1\Big),\\
V_{k\Delta}^{(2)} &:= \tfrac{2}{\Delta}\,\sigma\big(X_{k\Delta}\big)\Big[(W_{(k+1)\Delta} - W_{k\Delta}) \int_{k\Delta}^{(k+1)\Delta} \big(\sigma(X_s) - \sigma(X_{k\Delta})\big)\, dW_s\\
&\qquad - \int_{k\Delta}^{(k+1)\Delta} \big(\sigma(X_s) - \sigma(X_{k\Delta})\big)\, ds\Big],\\
V_{k\Delta}^{(3)} &:= \tfrac{2}{\Delta}\,\sigma\big(X_{k\Delta}\big)\Big[\int_{k\Delta}^{(k+1)\Delta} \big((k+1)\Delta - s\big)\, \sigma'(X_u)\, \sigma(X_u)\, dW_s\Big],\\
V_{k\Delta}^{(4)} &:= 2\, b\big(X_{k\Delta}\big)\, \sigma\big(X_{k\Delta}\big)\, \big(W_{(k+1)\Delta} - W_{k\Delta}\big).
\end{aligned}$$

The remainder terms are

$$R_{k\Delta}^{(1)} := \tfrac{1}{\Delta}\Big(\int_{k\Delta}^{(k+1)\Delta} b(X_s)\,ds\Big)^2, \quad R_{k\Delta}^{(2)} := \tfrac{1}{\Delta}\Big(\int_{k\Delta}^{(k+1)\Delta} \big(\sigma(X_s)-\sigma(X_{k\Delta})\big)\,dW_s\Big)^2,$$

$$R_{k\Delta}^{(3)} := \tfrac{2}{\Delta}\Big(\int_{k\Delta}^{(k+1)\Delta} \big(b(X_s)-b(X_{k\Delta})\big)\,ds\Big)\int_{k\Delta}^{(k+1)\Delta} \sigma(X_s)\,dW_s,$$

$$R_{k\Delta}^{(4)} := 2\,b\big(X_{k\Delta}\big)\int_{k\Delta}^{(k+1)\Delta} \big(\sigma(X_s)-\sigma(X_{k\Delta})\big)\,dW_s,$$

$$R_{k\Delta}^{(5)} := \tfrac{2}{\Delta}\,\sigma\big(X_{k\Delta}\big)\int_{k\Delta}^{(k+1)\Delta} \big((k+1)\Delta - s\big)\,\psi(X_s)\,ds,$$

where

$$\psi \;:=\; \sigma'\,b \;+\; \tfrac{1}{2}\,\sigma''\,\sigma^2. \tag{18}$$

In the arguments that follow, let

$$\overline{V}_{k\Delta} := V_{k\Delta}^{(2)} + V_{k\Delta}^{(3)} + V_{k\Delta}^{(4)},$$

contain all the secondary variance terms, while $V_{k\Delta}^{(1)}$ is considered the principal variance term.

Define the vectors

$$V := (V_\Delta, \ldots, V_{n\Delta})^T, \quad R := (R_\Delta, \ldots, R_{n\Delta})^T,$$

and the empirical processes

$$v_n(t) \;:=\; \langle t, V\rangle_n, \quad R_n(t) \;:=\; \langle t, R\rangle_n.$$

Then, for every $s,t \in S_m$, the identity

$$\widetilde{\gamma}_n(t) \;-\; \widetilde{\gamma}_n(s) \;=\; \|t-\sigma_A^2\|_n^2 \;-\; \|s-\sigma_A^2\|_n^2 \;-\; 2\,v_n\big(t-s\big) \;-\; 2\,R_n\big(t-s\big)$$

holds. Consequently,

$$\|\widehat{\sigma}_m^2-\sigma_A^2\|_n^2 \;\le\; \|t-\sigma_A^2\|_n^2 \;+\; 2\,v_n\big(\widehat{\sigma}_m^2-t\big) \;+\; 2\,R_n\big(\widehat{\sigma}_m^2-t\big). \tag{19}$$

Separate the noise into two groups with distinct behaviors:

$$v_n^{(1)}(t) \;:=\; \langle t, V^{(1)}\rangle_n, \quad \overline{v_n}(t) \;:=\; \langle t, V^{(2)}\rangle_n + \langle t, V^{(3)}\rangle_n + \langle t, V^{(4)}\rangle_n,$$

so that

$$v_n\big(\widehat{\sigma}_m^2-t\big) = v_n^{(1)}\big(\widehat{\sigma}_m^2-t\big) + \overline{v_n}\big(\widehat{\sigma}_m^2-t\big).$$

Then

$$v_n\big(\widehat{\sigma}_m^2-t\big) \;\le\; \sup_{\substack{t\in S_m\\ \|t\|_\pi=1}} \big(v_n^{(1)}(t)\big)^2\|\widehat{\sigma}_m^2-t\|_\pi^2 \;+\; \sup_{\substack{t\in S_m\\ \|t\|_\pi=1}} \big(\overline{v_n}(t)\big)^2\|\widehat{\sigma}_m^2-t\|_\pi^2. \tag{20}$$

For the residual terms, write:

$$R_n\big(\widehat{\sigma}_m^2-t\big) = \langle \widehat{\sigma}_m^2-t,\; R\rangle_n \;\le\; \|\widehat{\sigma}_m^2-t\|_n\,\|R\|_n. \tag{21}$$

Substitute (20) and (21) into (19), and apply $2\,ab \le \frac{1}{C}\,a^2 + C\,b^2$ with $C=8$, leading to

$$\begin{aligned}\|\widehat{\sigma}_m^2-\sigma_A^2\|_n^2 \le \|t-\sigma_A^2\|_n^2 \;+\; 8\sup_{\substack{t\in S_m\\ \|t\|_\pi=1}} \big(v_n^{(1)}(t)\big)^2 \;+\; \tfrac{1}{8}\,\|\widehat{\sigma}_m^2-t\|_\pi^2 \;+\; \tfrac{1}{8}\,\|\widehat{\sigma}_m^2-t\|_n^2 \\ +\; 8\sup_{\substack{t\in S_m\\ \|t\|_\pi=1}} \big(\overline{v_n}(t)\big)^2 \;+\; 8\,\|R\|_n^2.\end{aligned} \tag{22}$$

Inequality $\|x+y\|_n^2 \le 2\,\|x\|_n^2 + 2\,\|y\|_n^2$ implies

$$\|\widehat{\sigma}_m - t\|_n^2 \;\le\; 2\,\|t-\sigma_A^2\|_n^2 \;+\; 2\,\|\widehat{\sigma}_m^2-\sigma_A^2\|_n^2.$$

On Ω_m, for functions in S_m, there is also $\|\cdot\|_\pi \le 2\,\|\cdot\|_n$. Thus,

$$\tfrac{1}{8}\,\|\widehat{\sigma}_m^2-t\|_\pi^2 \;+\; \tfrac{1}{8}\,\|\widehat{\sigma}_m^2-t\|_n^2 \;\le\; \Big(\tfrac{1}{4}+\tfrac{1}{8}\Big)\,\|\widehat{\sigma}_m^2-t\|_n^2 \;\le\; \tfrac{3}{4}\,\|t-\sigma_A^2\|_n^2 \;+\; \tfrac{3}{4}\,\|\widehat{\sigma}_m^2-\sigma_A^2\|_n^2.$$

Hence, for any $t \in S_m$,

$$\|\widehat{\sigma}_m^2-\sigma_A^2\|_n^2 \;\le\; \tfrac{7}{4}\,\|t-\sigma_A^2\|_n^2 \;+\; \tfrac{3}{4}\,\|\widehat{\sigma}_m^2-\sigma_A^2\|_n^2 \;+\; 8\sup_{\substack{t\in S_m\\ \|t\|_\pi=1}} \big(v_n^{(1)}(t)\big)^2 \;+\; 8\sup_{\substack{t\in S_m\\ \|t\|_\pi=1}} \big(\overline{v_n}(t)\big)^2 \;+\; \tfrac{8}{n}\,\|R\|_{2,n}^2.$$

Rearranging the last term on the left-hand side yields, on $\Lambda_m \cap \Omega_m$,

$$\|\widehat{\sigma}_m^2 - \sigma_A^2\|_n^2 \le 7\,\|t - \sigma_A^2\|_n^2 \;+\; 32 \sup_{\substack{t\in S_m\\ \|t\|_\pi=1}} \big(v_n^{(1)}(t)\big)^2 \;+\; 32 \sup_{\substack{t\in S_m\\ \|t\|_\pi=1}} \big(\overline{v_n}(t)\big)^2 \;+\; \tfrac{64}{n}\,\|R\|_{2,n}^2. \tag{23}$$

Lemmas for bounding the noise and remainder terms.

Lemma 5.2. *Under the assumptions of Proposition 1, for every $k \in \mathbb{N}$,*

$$\mathbb{E}\big[\big(V_{k\Delta}^{(i)}\big)^2\big] \lesssim \Delta, \quad \mathbb{E}\big[\big(V_{k\Delta}^{(i)}\big)^4\big] \lesssim \Delta^2, \quad i \in \{2,3,4\}, \tag{24}$$

$$\mathbb{E}\big[\big(R_{k\Delta}^{(i)}\big)^2\big] \lesssim \Delta^2, \quad \mathbb{E}\big[\big(R_{k\Delta}^{(i)}\big)^4\big] \lesssim \Delta^4, \quad i \in \{1,2,3,4,5\}. \tag{25}$$

Lemma 5.3. *Under the assumptions of Proposition 1,*

$$\mathbb{E}\Big[\sup_{\substack{t\in S_m\\ \|t\|_\pi=1}} \big(v_n^{(1)}(t)\big)^2\Big] = 2\,\frac{\mathrm{Tr}\Big[\Psi_m^{-1/2}\,\Psi_{m,\sigma^4}\,\Psi_m^{-1/2}\Big]}{n}, \tag{26}$$

$$\mathbb{E}\Big[\sup_{\substack{t\in S_m\\ \|t\|_\pi=1}} \big(\overline{v_n}(t)\big)^2\Big] \lesssim \Delta^2. \tag{27}$$

Combining (26), (27), (25), and (23), and enforcing (10) for m, leads to

$$\mathbb{E}\Big[\|\widehat{\sigma}_m^2 - \sigma_A^2\|_n^2\,\mathbb{1}_{\Lambda_m\cap\Omega_m}\Big] \;\le\; 7\,\mathbb{E}\big[\|t-\sigma_A^2\|_n^2\big] \;+\; 64\,\frac{\mathrm{Tr}\big[\Psi_m^{-1/2}\,\Psi_{m,\sigma^4}\,\Psi_m^{-1/2}\big]}{n} \;+\; c\,\Delta^2.$$

This completes the second step of the proof for $\mathbb{E}\big[\|\widehat{\sigma}_m^2 - \sigma_A^2\|_n^2\,\mathbb{1}_{\Lambda_m\cap\Omega_m}\big]$.

Step 3. Bound on $\mathbb{E}\big[\|\widehat{\sigma}_m^2 - \sigma_A^2\|_n^2\,\mathbb{1}_{\Lambda_m\cap(\Omega_m)^c}\big]$.

Let Π_m denote the orthogonal projection onto the subspace S_m. Similarly to (7), define

$$\Pi_m\sigma_A^2 \;=\; \sum_{j=0}^{m-1} \tilde{c}_j^{(m)}\,\varphi_j, \quad \text{where} \quad \tilde{c}^{(m)} \;=\; \tfrac{1}{n}\,\widehat{\Psi}_m^{-1}\,\widehat{\Phi}_m^T\,\sigma^2,$$

and $\sigma^2 := (\sigma(X_\Delta)^2, \dots, \sigma(X_{n\Delta})^2)^T$. Hence,

$$\widehat{\sigma}_m^2 \;-\; \Pi_m\sigma_A^2 \;=\; \sum_{j=0}^{m-1} \tilde{\tilde{c}}_j^{(m)}\,\varphi_j, \quad \text{where} \quad \tilde{\tilde{c}}^{(m)} \;=\; \tfrac{1}{n}\,\widehat{\Psi}_m^{-1}\,\widehat{\Phi}_m^T\,E,$$

and $E := (V_\Delta + R_\Delta, \dots, V_{n\Delta} + R_{n\Delta})^T$. Thus,

$$\|\widehat{\sigma}_m^2 - \Pi_m\sigma_A^2\|_n^2 \;=\; \frac{1}{n}\sum_{i=1}^{n}\Big(\sum_{j=0}^{m-1} \tilde{\tilde{c}}_j^{(m)}\,\varphi_j(X_{i\Delta})\Big)^2 \;=\; \big(\tilde{\tilde{c}}^{(m)}\big)^T\,\widehat{\Psi}_m\,\tilde{\tilde{c}}^{(m)}.$$

Consequently,

$$\|\widehat{\sigma}_m^2 - \Pi_m\sigma_A^2\|_n^2 \;=\; \tfrac{1}{n^2}\,E^T\,\widehat{\Phi}_m\,\widehat{\Psi}_m^{-1}\,\widehat{\Phi}_m^T\,E \;\le\; \|\widehat{\Psi}_m^{-1}\|_{\mathrm{op}}\,\Big\|\tfrac{1}{n}\,\widehat{\Phi}_m^T\,E\Big\|_{2,m}^2. \tag{28}$$

The following lemma provides a useful bound.

Lemma 5.4. *Under the assumptions of Proposition 1, there is a constant C such that*

$$\mathbb{E}\Big[\Big\|\tfrac{1}{n}\,(\widehat{\Phi}_m)^T\,(V+R)\Big\|_{2,m}^4\Big] \;\le\; C\,m\,L(m)^2\,\Big(\tfrac{1}{n^2} \;+\; \Delta^4\Big).$$

Applying Lemma 5.4 and the definition of Λ_m in (28) leads to

$$\begin{aligned}\mathbb{E}\Big[\|\widehat{\sigma}_m^2 - \Pi_m\sigma_A^2\|_n^2\,\mathbb{1}_{\Lambda_m\cap(\Omega_m)^c}\Big] &\lesssim \frac{n\Delta}{L(m)\,\log^2(n\Delta)}\,\mathbb{E}^{1/2}\Big[\Big\|\tfrac{1}{n}\,(\widehat{\Phi}_m)^T\,E\Big\|_{2,m}^4\Big]\,\mathbb{P}\big((\Omega_m)^c\big)^{1/2}\\ &\lesssim \frac{n\Delta}{\log^2(n\Delta)}\,(n\Delta)^{1/2}\,\sqrt{\tfrac{1}{n^2}+\Delta^4}\,\tfrac{1}{(n\Delta)^3} \;\lesssim\; \frac{1}{n} \;+\; \Delta^2. \end{aligned} \tag{29}$$

Combining the last display with a simple decomposition yields

$$\mathbb{E}\Big[\|\widehat{\sigma}_m^2 - \sigma_A^2\|_n^2\,\mathbb{1}_{\Lambda_m\cap(\Omega_m)^c}\Big] \;\le\; \mathbb{E}\Big[\|\widehat{\sigma}_m^2 - \Pi_m\sigma_A^2\|_n^2\,\mathbb{1}_{\Lambda_m\cap(\Omega_m)^c}\Big] \;+\; \mathbb{E}\Big[\|\Pi_m\sigma_A^2 - \sigma_A^2\|_n^2\,\mathbb{1}_{\Lambda_m\cap(\Omega_m)^c}\Big].$$

Note that

$$\mathbb{E}\Big[\|\Pi_m\sigma_A^2-\sigma_A^2\|_n^2\,\mathbb{1}_{\Lambda_m\cap(\Omega_m)^c}\Big]\;\le\;\mathbb{E}\Big[\|\sigma_A^2\|_n^2\,\mathbb{1}_{(\Omega_m)^c}\Big].$$

Hence,

$$\mathbb{E}\Big[\|\widehat{\sigma}_m^2-\sigma_A^2\|_n^2\,\mathbb{1}_{\Lambda_m\cap(\Omega_m)^c}\Big]\;\lesssim\;\frac{1}{n}\;+\;\Delta^2.$$

This completes the derivation of the desired bound for the third term.

5.2. **Proof of Lemma 5.2.**
Use standard inequalities, in particular the Cauchy–Schwarz (C–S) and Burkholder–Davis–Gundy (BDG) inequalities. Only the bounds for the fourth-order moments are shown, because the bounds for the second-order moments follow from these results via C–S: for any random variable X, it holds that $\mathbb{E}[X^2]\le\mathbb{E}[X^4]^{1/2}$.

For simplicity, set

$$\begin{aligned}
&G=\int_{k\Delta}^{(k+1)\Delta}b(X_s)\,ds,\quad H=\int_{k\Delta}^{(k+1)\Delta}\sigma(X_s)\,dW_s,\\
&I=\int_{k\Delta}^{(k+1)\Delta}\big(\sigma(X_s)-\sigma(X_{k\Delta})\big)\,dW_s,\quad J=\int_{k\Delta}^{(k+1)\Delta}\big(\sigma(X_s)-\sigma(X_{k\Delta})\big)\,ds,\\
&K=\int_{k\Delta}^{(k+1)\Delta}\big((k+1)\Delta-s\big)\sigma'(X_s)\sigma(X_s)\,dW_s,\quad L=\int_{k\Delta}^{(k+1)\Delta}\big((k+1)\Delta-s\big)\psi(X_s)\,ds,\\
&M=\int_{k\Delta}^{(k+1)\Delta}\big(b(X_s)-b(X_{k\Delta})\big)\,ds,\quad N=W_{(k+1)\Delta}-W_{k\Delta}.
\end{aligned}$$

Step 1: Bounding the eighth-order moments of G,H,I,J,K,L,M,N.

Bound for $\mathbb{E}[G^8]$.

$$\begin{aligned}
\mathbb{E}[G^8]&=\mathbb{E}\Big[\Big(\int_{k\Delta}^{(k+1)\Delta}b(X_s)\,ds\Big)^8\Big]\;\le\;\Delta^4\,\mathbb{E}\Big[\Big(\int_{k\Delta}^{(k+1)\Delta}b^2(X_s)\,ds\Big)^4\Big]\\
&\le\Delta^6\,\mathbb{E}\Big[\Big(\int_{k\Delta}^{(k+1)\Delta}b^4(X_s)\,ds\Big)^2\Big]\;\le\;\Delta^7\,\mathbb{E}\Big[\int_{k\Delta}^{(k+1)\Delta}b^8(X_s)\,ds\Big]\\
&=\Delta^7\int_{k\Delta}^{(k+1)\Delta}\mathbb{E}[b^8(X_0)]\,ds\;\lesssim\;\Delta^8.
\end{aligned}$$

The estimate uses three successive applications of C–S, stationarity of $(X_t)_{t\ge0}$, Assumption 1(i), and the first part of (2).

Bound for $\mathbb{E}[H^8]$.

$$\begin{aligned}
\mathbb{E}[H^8]&=\mathbb{E}\Big[\Big(\int_{k\Delta}^{(k+1)\Delta}\sigma(X_s)\,dW_s\Big)^8\Big]\;\lesssim\;\mathbb{E}\Big[\Big(\int_{k\Delta}^{(k+1)\Delta}\sigma^2(X_s)\,ds\Big)^4\Big]\\
&\lesssim\Delta^3\,\mathbb{E}\Big[\int_{k\Delta}^{(k+1)\Delta}\sigma^8(X_s)\,ds\Big]\;=\;\Delta^3\int_{k\Delta}^{(k+1)\Delta}\mathbb{E}[\sigma^8(X_0)]\,ds\;\lesssim\;\Delta^4.
\end{aligned}$$

The initial step applies the BDG inequality, followed by C–S and arguments similar to those used for G.

Bound for $\mathbb{E}[I^8]$.

$$\begin{aligned}
\mathbb{E}[I^8]&\lesssim\mathbb{E}\Big[\Big(\int_{k\Delta}^{(k+1)\Delta}\big(\sigma(X_s)-\sigma(X_{k\Delta})\big)^2\,ds\Big)^4\Big]\\
&\lesssim\Delta^3\,\mathbb{E}\Big[\int_{k\Delta}^{(k+1)\Delta}\big(\sigma(X_s)-\sigma(X_{k\Delta})\big)^8\,ds\Big]\\
&\lesssim\Delta^3\int_{k\Delta}^{(k+1)\Delta}\mathbb{E}\Big[\sup_{k\Delta\le t\le(k+1)\Delta}\big|\sigma(X_t)-\sigma(X_{k\Delta})\big|^8\Big]\,ds\;\lesssim\;\Delta^3\int_{k\Delta}^{(k+1)\Delta}\Delta^4\,ds\;\lesssim\;\Delta^8.
\end{aligned}$$

The BDG and C–S inequalities are used, along with (3).

Bound for $\mathbb{E}[J^8]$.

$$\mathbb{E}[J^8] \lesssim \Delta^7\,\mathbb{E}\Big[\int_{k\Delta}^{(k+1)\Delta} \big(\sigma(X_s) - \sigma(X_{k\Delta})\big)^8\,ds\Big] \lesssim \Delta^{12}.$$

The same arguments as for I are employed, except that the BDG inequality is not needed.

Bound for $\mathbb{E}[K^8]$.

$$\begin{aligned}\mathbb{E}[K^8] &\lesssim \mathbb{E}\Big[\Big(\int_{k\Delta}^{(k+1)\Delta} \big((k+1)\Delta - s\big)^2\,\sigma'(X_s)^2\,\sigma(X_s)^2\,ds\Big)^4\Big]\\ &\lesssim \Delta^3\,\mathbb{E}\Big[\int_{k\Delta}^{(k+1)\Delta} \big((k+1)\Delta - s\big)^8\,\sigma'(X_s)^8\,\sigma(X_s)^8\,ds\Big]\\ &\lesssim \Delta^3 \int_{k\Delta}^{(k+1)\Delta} \Delta^8\,\mathbb{E}\big[\sigma'(X_0)^8\,\sigma(X_0)^8\big]\,ds \lesssim \Delta^{12}.\end{aligned}$$

The BDG and C–S inequalities are used again, along with stationarity, Assumption 2(i), and the first part of (2).

Bound for $\mathbb{E}[L^8]$.

$$\mathbb{E}[L^8] \lesssim \Delta^7\,\mathbb{E}\Big[\int_{k\Delta}^{(k+1)\Delta} \Delta^8\,\psi(X_s)\,ds\Big] \lesssim \Delta^{16}.$$

The same arguments as for J apply, together with Assumption 1(ii) to bound the function ψ.

Bound for $\mathbb{E}[M^8]$.

$$\mathbb{E}[M^8] \lesssim \Delta^{12},$$

which follows the same reasoning as for J.

Bound for $\mathbb{E}[N^8]$. It is well known that $\mathbb{E}[N^8] = 105\,\Delta^4$ because N is a Gaussian increment of the Brownian motion over an interval of length Δ.

Step 2: Combining the estimates to bound the terms in the lemma.

- $\mathbb{E}\big[(V_{k\Delta}^{(2)})^4\big] = \dfrac{16}{\Delta^4}\,\mathbb{E}\big[(NI - J)^4\big] \lesssim \dfrac{1}{\Delta^4}\,\mathbb{E}[\sigma^8(X_{k\Delta})]^{\frac12}\,\mathbb{E}[(NI-J)^8]^{\frac12} \lesssim \dfrac{1}{\Delta^4}\Big(\mathbb{E}[N^8]\mathbb{E}[I^8] + \mathbb{E}[J^8]\Big)^{\frac12} \lesssim \Delta^2$.
- $\mathbb{E}\big[(V_{k\Delta}^{(3)})^4\big] \lesssim \dfrac{1}{\Delta}\,\mathbb{E}[\sigma^8(X_{k\Delta})]^{\frac12}\,\mathbb{E}[K^8]^{\frac12} \lesssim \Delta^2$.
- $\mathbb{E}\big[(V_{k\Delta}^{(4)})^4\big] \le \mathbb{E}[b^8(X_{k\Delta})\,\sigma^8(X_{k\Delta})]^{\frac12}\,\mathbb{E}[N^8]^{\frac12}$.
- $\mathbb{E}\big[(R_{k\Delta}^{(1)})^4\big] = \dfrac{1}{\Delta^4}\,\mathbb{E}[G^8] \lesssim \Delta^4$.
- $\mathbb{E}\big[(R_{k\Delta}^{(2)})^4\big] = \dfrac{1}{\Delta^4}\,\mathbb{E}[I^8] \lesssim \Delta^4$.
- $\mathbb{E}\big[(R_{k\Delta}^{(3)})^4\big] = \dfrac{16}{\Delta^4}\,\mathbb{E}[M^4\,H^4] \lesssim \dfrac{1}{\Delta^4}\,\mathbb{E}[M^8]^{\frac12}\,\mathbb{E}[H^8]^{\frac12} \lesssim \Delta^4$.
- $\mathbb{E}\big[(R_{k\Delta}^{(4)})^4\big] = 16\,\mathbb{E}\big[b^4(X_{k\Delta})\,I^4\big] \le 16\,\mathbb{E}[b^8(X_{k\Delta})]^{\frac12}\,\mathbb{E}[I^8]^{\frac12} \lesssim \Delta^4$.
- $\mathbb{E}\big[(R_{k\Delta}^{(5)})^4\big] \lesssim \dfrac{1}{\Delta^4}\,\mathbb{E}[L^8]^{\frac12} \lesssim \Delta^4$.

Each estimate follows directly from the established eighth-moment bounds of the variables G, H, I, J, K, L, M, and N presented above. In particular, the estimate for $\mathbb{E}\left[\left(V_{k\Delta}^{(2)}\right)^4\right]$ uses the independent increments property of the Brownian motion, which allows us to separate the terms involving N and I. Consequently, all required bounds are satisfied, thus completing the proof of the lemma 5.2.

5.3. **Proof of Lemma 5.3.**

Recall from (5) that

$$U_{k\Delta} \;=\; \sigma\big(X_{k\Delta}\big)^2 \;+\; V_{k\Delta}^{(1)} \;+\; \overline{V}_{k\Delta} \;+\; R_{k\Delta},$$

where

$$V_{k\Delta}^{(1)} \;=\; \sigma^2\big(X_{k\Delta}\big)\left(\tfrac{(W_{(k+1)\Delta} - W_{k\Delta})^2}{\Delta} - 1\right), \quad \overline{V}_{k\Delta} \;=\; V_{k\Delta}^{(2)} + V_{k\Delta}^{(3)} + V_{k\Delta}^{(4)}.$$

In particular,

$$\mathbb{E}\big[V_{u\Delta}^{(1)} \mid \mathcal{F}_{u\Delta}\big] \;=\; \sigma^2\big(X_{u\Delta}\big)\cdot 0 \;=\; 0.$$

Define the vectors

$$V^{(1)} \;:=\; \big(V_{\Delta}^{(1)},\dots,V_{n\Delta}^{(1)}\big)^T, \quad \overline{V} \;:=\; \big(\overline{V}_{\Delta},\dots,\overline{V}_{n\Delta}\big)^T.$$

Then for any function t,

$$v_n^{(1)}(t) \;=\; \langle t,\, V^{(1)}\rangle_n \;=\; \frac{1}{n}\sum_{k=1}^{n} t\big(X_{k\Delta}\big)\, V_{k\Delta}^{(1)}.$$

Proof of (26). We begin by introducing

$$V_m \;:=\; \sup_{\substack{t\in S_m\\ \|t\|_\pi=1}} \Big(v_n^{(1)}(t)\Big)^2.$$

Our aim is to compute $\mathbb{E}\big[V_m\big]$. Since any $t \in S_m$ can be written as $t(x) = \sum_{j=0}^{m-1} a_j\, \varphi_j(x)$ and $\|t\|_\pi^2 = \sum_{j,k} a_j\, a_k\, [\Psi_m]_{j,k}$, it follows that

$$\|t\|_\pi = 1 \quad\Longleftrightarrow\quad \|\Psi_m^{1/2} a\|_{2,m} = 1.$$

Hence

$$V_m \;=\; \sup_{\|a\|_{2,m}=1} \Big\langle V^{(1)},\, \sum_{j=0}^{m-1} \big[\Psi_m^{-1/2}\big]_{j,\cdot}\, \varphi_j \Big\rangle_n^2 \;=\; \sum_{k=0}^{m-1} \Big\langle V^{(1)},\, \sum_{j=0}^{m-1} \big[\Psi_m^{-1/2}\big]_{j,k}\, \varphi_j \Big\rangle_n^2,$$

by a standard Cauchy–Schwarz argument (there is indeed a maximizing t). Denote

$$f_k(x) \;=\; \sum_{j=0}^{m-1} [\Psi_m^{-1/2}]_{j,k}\, \varphi_j(x).$$

Then

$$\Big\langle V^{(1)},\, \sum_{j=0}^{m-1} \big[\Psi_m^{-1/2}\big]_{j,k}\, \varphi_j \Big\rangle_n \;=\; \frac{1}{n}\sum_{u=1}^{n} V_{u\Delta}^{(1)}\, f_k\big(X_{u\Delta}\big).$$

Taking expectation and exploiting the fact that $\mathbb{E}[V_{u\Delta}^{(1)} \mid \mathcal{F}_{u\Delta}] = 0$, the cross-terms $u \neq v$ vanish:

$$\mathbb{E}\Big[\Big\langle V^{(1)},\, \sum_{j=0}^{m-1} \big[\Psi_m^{-1/2}\big]_{j,k}\, \varphi_j \Big\rangle_n^2\Big] \;=\; \frac{1}{n^2}\sum_{u=1}^{n} \mathbb{E}\big[\big(V_{u\Delta}^{(1)}\big)^2\, f_k(X_{u\Delta})^2\big] \;=\; \frac{1}{n}\,\mathbb{E}\big[\sigma^4(X_0)\, f_k(X_0)^2\big] \times 2,$$

since $\mathrm{Var}[(W_{(u+1)\Delta} - W_{u\Delta})^2/\Delta - 1] = 2$. Next, expand $f_k(X_0)^2$:

$$f_k(X_0)^2 \;=\; \sum_{j,\ell=0}^{m-1} [\Psi_m^{-1/2}]_{j,k}\, [\Psi_m^{-1/2}]_{\ell,k}\, \varphi_j(X_0)\, \varphi_\ell(X_0).$$

Thus,

$$\mathbb{E}\big[f_k(X_0)^2\, \sigma^4(X_0)\big] \;=\; \sum_{j,\ell=0}^{m-1} [\Psi_m^{-1/2}]_{j,k}\, [\Psi_m^{-1/2}]_{\ell,k}\, \mathbb{E}\big[\varphi_j(X_0)\, \varphi_\ell(X_0)\, \sigma^4(X_0)\big] \;=\; \sum_{j,\ell=0}^{m-1} [\Psi_m^{-1/2}]_{j,k}\, [\Psi_m^{-1/2}]_{\ell,k}\, [\Psi_{m,\sigma^4}]_{j,\ell},$$

where Ψ_{m,σ^4} is the matrix whose (j,ℓ)-th entry is $\mathbb{E}[\varphi_j(X_0)\, \varphi_\ell(X_0)\, \sigma^4(X_0)]$. Finally, summing over $k = 0,\dots,m-1$ gives

$$\mathrm{Tr}\big[\Psi_m^{-1/2}\, \Psi_{m,\sigma^4}\, \Psi_m^{-1/2}\big].$$

Hence

$$\mathbb{E}[V_m] \;=\; 2\,\frac{\mathrm{Tr}\big[\Psi_m^{-1/2}\, \Psi_{m,\sigma^4}\, \Psi_m^{-1/2}\big]}{n},$$

which is precisely (26).

Proof of (27). We now prove

$$\mathbb{E}\Big[\sup_{\substack{t\in S_m\\ \|t\|_\pi=1}} \big(\overline{v_n}(t)\big)^2\Big] \lesssim \Delta^2.$$

Recall $\overline{v_n}(t) = \langle t, \overline{V}\rangle_n$. By definition,

$$\overline{V}_{k\Delta} = V^{(2)}_{k\Delta} + V^{(3)}_{k\Delta} + V^{(4)}_{k\Delta},$$

where each $V^{(i)}_{k\Delta}$ has zero conditional mean given $\mathcal{F}_{k\Delta}$. Let

$$\overline{v_n}(t) = \frac{1}{n}\sum_{k=1}^{n} t(X_{k\Delta})\,\overline{V}_{k\Delta}.$$

We decompose

$$\mathbb{E}\Big[\big\langle \overline{V},\ \sum_{j=0}^{m-1}[\Psi_m^{-1/2}]_{j,k}\,\varphi_j\big\rangle_n^2\Big] = \frac{1}{n^2}\sum_{u,v=1}^{n}\mathbb{E}\Big[\overline{V}_{u\Delta}\,\overline{V}_{v\Delta}\,f_k(X_{u\Delta})\,f_k(X_{v\Delta})\Big].$$

As in the proof of (26), the cross terms $u \neq v$ vanish under expectation (due to the martingale property and mixing), so only the diagonal remains. Also, using $\|\widehat{\Psi}_m^{-1}\|_{\mathrm{op}}$ and $L(m) := \sup_{x\in A}\sum_{j=0}^{m-1}\varphi_j(x)^2$, we obtain

$$\sum_{k=0}^{m-1}\sum_{j,\ell=0}^{m-1}[\Psi_m^{-1/2}]_{j,k}\,[\Psi_m^{-1/2}]_{\ell,k} \le \|\Psi_m^{-1}\|_{\mathrm{op}}\,m \quad\text{and}\quad \max_{0\le j\le m-1}\|\varphi_j\|_\infty^2 \le L(m).$$

By Lemma 5.2 (or a direct calculation), $\mathbb{E}[\overline{V}^2_{k\Delta}] \lesssim \Delta$. Hence,

$$\mathbb{E}\Big[\sup_{\|t\|_\pi=1}\overline{v_n}(t)^2\Big] \lesssim \frac{m\,\|\Psi_m^{-1}\|_{\mathrm{op}}\,\Delta}{n} \le \frac{c_\varphi^2\,m\,\|\Psi_m^{-1}\|_{\mathrm{op}}}{n}\,\Delta \lesssim \Delta^2,$$

as soon as $m\,\|\Psi_m^{-1}\|_{\mathrm{op}} \lesssim (n\Delta)/\log^2(n\Delta)$. This proves (27).

5.4. **Proof of Lemma 5.4.**
Recall that $E := V + R$. A Cauchy–Schwarz argument yields

$$\Big\|\tfrac{1}{n}\,\widehat{\Phi}_m^T E\Big\|_{2,m}^4 = \tfrac{1}{n^4}\Big(\sum_{j=0}^{m-1}\Big(\sum_{i=1}^{n}\varphi_j(X_{i\Delta})\,E_{i\Delta}\Big)^2\Big)^2 \le \tfrac{m}{n^4}\sum_{j=0}^{m-1}\Big(\sum_{i=1}^{n}\varphi_j(X_{i\Delta})\,E_{i\Delta}\Big)^4.$$

Next, apply $(a+b)^4 \le 8\,(a^4+b^4)$ to separate the summation into two parts:

$$\Big\|\tfrac{1}{n}\,\widehat{\Phi}_m^T E\Big\|_{2,m}^4 \lesssim \tfrac{m}{n^4}\sum_{j=0}^{m-1}\Big[\Big(\sum_{i=1}^{n}\varphi_j(X_{i\Delta})\,V_{i\Delta}\Big)^4 + \Big(\sum_{i=1}^{n}\varphi_j(X_{i\Delta})\,R_{i\Delta}\Big)^4\Big].$$

Define

$$T_V := \tfrac{m}{n^4}\sum_{j=0}^{m-1}\Big(\sum_{i=1}^{n}\varphi_j(X_{i\Delta})\,V_{i\Delta}\Big)^4 \quad\text{and}\quad T_R := \tfrac{m}{n^4}\sum_{j=0}^{m-1}\Big(\sum_{i=1}^{n}\varphi_j(X_{i\Delta})\,R_{i\Delta}\Big)^4,$$

so

$$\Big\|\tfrac{1}{n}\,\widehat{\Phi}_m^T E\Big\|_{2,m}^4 \lesssim T_V + T_R.$$

Term involving $\boldsymbol{R}$. A double Cauchy–Schwarz argument implies

$$\Big(\sum_{i=1}^{n}\varphi_j(X_{i\Delta})\,R_{i\Delta}\Big)^4 \le \Big(n\sum_{i=1}^{n}\varphi_j(X_{i\Delta})^2\,R_{i\Delta}^2\Big)^2 \le n^3\sum_{i=1}^{n}\varphi_j(X_{i\Delta})^4\,R_{i\Delta}^4.$$

By Lemma 5.2, $\mathbb{E}[R_0^4] \lesssim \Delta^4$. Also, $\sup_x\sum_{j=0}^{m-1}\varphi_j(x)^4 \le L(m)^2$. Hence, taking expectation gives

$$\mathbb{E}[T_R] = \mathbb{E}\Big[\tfrac{m}{n^4}\sum_{j=0}^{m-1}\Big(\sum_{i=1}^{n}\varphi_j(X_{i\Delta})\,R_{i\Delta}\Big)^4\Big] \lesssim m\,L(m)^2\,\Delta^4.$$

Term involving $\boldsymbol{V}$. Since $\{\varphi_j(X_{i\Delta})\,V_{i\Delta}\}_i$ are uncorrelated and have mean zero,

$$\mathbb{E}\Big[\Big(\sum_{i=1}^{n}\varphi_j(X_{i\Delta})\,V_{i\Delta}\Big)^4\Big] \;\lesssim\; n\,\mathbb{E}\big[\varphi_j(X_0)^4\,V_0^4\big] \;+\; n^2\left(\mathbb{E}[\varphi_j(X_0)^2\,V_0^2]\right)^2.$$

It is straightforward to check that V_0 has bounded moments and $\sup_x \sum_{j=0}^{m-1}\varphi_j(x)^4 \le L(m)^2$. Hence,

$$\mathbb{E}[T_V] \;=\; \mathbb{E}\Big[\tfrac{m}{n^4}\sum_{j=0}^{m-1}\Big(\sum_{i=1}^{n}\varphi_j(X_{i\Delta})\,V_{i\Delta}\Big)^4\Big] \;\lesssim\; \frac{m\,L(m)^2}{n^2}.$$

Combining the above bounds completes the argument:

$$\mathbb{E}\Big[\Big\|\tfrac{1}{n}\,\widehat{\Phi}_m^T\,(V+R)\Big\|_{2,m}^4\Big] \;\lesssim\; m\,L(m)^2\Big(\tfrac{1}{n^2}+\Delta^4\Big),$$

which proves Lemma 5.4.

5.5. **Proof of Theorem 3.1.**

The strategy follows that of Theorem 3.1 in [7], requiring that X_t possesses finite high-order moments. This requirement does not impose stricter conditions on Model 1, thanks to the geometric β-mixing condition stated in (2).

Definitions. Let $\widehat{M}_{n\Delta}$ be the maximal element of $\widehat{\mathcal{M}}_{n\Delta}$ (see (15)). Let $M_{n\Delta}$ be the maximal element of $\mathcal{M}_{n\Delta}$, and let $M^+_{n\Delta}$ be the maximal element of the set

$$\mathcal{M}^+_{n\Delta} \;:=\; \Big\{m\in\mathbb{N}:\; c_\varphi^2\,m\,\big(\|\Psi_m^{-1}\|_{\mathrm{op}}^2\,\vee\,1\big)\;\le\;4\,\mathfrak{d}\,\frac{n\Delta}{\log^2(n\Delta)}\Big\}.$$

Here, $\mathfrak{d}$ is a positive constant defined in (14). Define

$$\Xi_{n\Delta} \;:=\; \big\{\,\mathcal{M}_{n\Delta}\;\subset\;\widehat{\mathcal{M}}_{n\Delta}\;\subset\;\mathcal{M}^+_{n\Delta}\big\},\quad \Omega_{n\Delta}\;:=\;\bigcap_{m\in\mathcal{M}^+_{n\Delta}}\Omega_m$$

and

$$p(m,m') := \kappa_0\,c_\varphi^2\,\mathbb{E}\big[\sigma^4(X_0)\big]\,\frac{(m\vee m')\,\|\Psi^{-1}_{m\vee m'}\|_{\mathrm{op}}}{n}.$$

$$\widehat{\mathrm{pen}}(m) \;=\; \kappa\,c_\varphi^2\,\mathbb{E}\big[\sigma^4(X_0)\big]\,\frac{m\,\|\widehat{\Psi}_m^{-1}\|_{\mathrm{op}}}{n}.$$

Lemma 5.5 below ensures that $\Xi_{n\Delta}$ and $\Omega_{n\Delta}$ each occur with high probability.

Lemma 5.5. *Under the assumptions of Theorem 3.1, there exist positive constants c, c' (independent of n and Δ) such that*

$$\mathbb{P}\big(\Omega^c_{n\Delta}\big)\;\le\;\frac{c}{(n\Delta)^5},\quad \mathbb{P}\big(\Xi^c_{n\Delta}\big)\;\le\;\frac{c'}{(n\Delta)^5}.$$

The proof of Lemma 5.5 is analogous to that of Lemma 6.4 in [7], except for a larger constant C_0 in the definition of $\mathcal{M}^+_{n\Delta}$, which modifies the probability rate to $(n\Delta)^{-5}$.

We write

$$\mathbb{E}\Big[\|\widehat{\sigma}^2_{\widehat{m}}-\sigma_A^2\|_n^2\Big] \;=\; \mathbb{E}\Big[\|\widehat{\sigma}^2_{\widehat{m}}-\sigma_A^2\|_n^2\,\mathbb{1}_{\Xi_{n\Delta}\cap\Omega_{n\Delta}}\Big] \;+\; \mathbb{E}\Big[\|\widehat{\sigma}^2_{\widehat{m}}-\sigma_A^2\|_n^2\,\mathbb{1}_{(\Xi_{n\Delta}\cap\Omega_{n\Delta})^c}\Big].$$

The proof thus splits into bounding these two terms on the right-hand side.

1. Bound on $\mathbb{E}\big[\|\widehat{\sigma}^2_{\widehat{m}}-\sigma_A^2\|_n^2\,\mathbb{1}_{(\Xi_{n\Delta}\cap\Omega_{n\Delta})^c}\big]$.

Decompose this expectation in two parts:

$$\mathbb{E}\Big[\|\widehat{\sigma}^2_{\widehat{m}}-\sigma_A^2\|_n^2\,\mathbb{1}_{(\Xi_{n\Delta}\cap\Omega_{n\Delta})^c}\Big] \;\le\; 2\,\mathbb{E}\Big[\|\widehat{\sigma}^2_{\widehat{m}}-\Pi_{\widehat{m}}\sigma_A^2\|_n^2\,\mathbb{1}_{(\Xi_{n\Delta}\cap\Omega_{n\Delta})^c}\Big] \;+\; 2\,\mathbb{E}\Big[\|\sigma_A^2\|_n^2\,\mathbb{1}_{(\Xi_{n\Delta}\cap\Omega_{n\Delta})^c}\Big].$$

First, note that

$$\|\widehat{\sigma}^2_{\widehat{m}}-\sigma_A^2\|_n^2 \;\le\; 2\,\|\widehat{\sigma}^2_{\widehat{m}}-\Pi_{\widehat{m}}\sigma_A^2\|_n^2 \;+\; 2\,\|\sigma_A^2\|_n^2,$$

since $\mathrm{Id}_{\widehat{m}}-\Pi_{\widehat{m}}$ is a projector of norm at most 1.

(i) Bounding $\|\widehat{\sigma}^2_{\widehat{m}}-\Pi_{\widehat{m}}\sigma_A^2\|_n^2$.

An argument similar to (28) shows

$$\|\widehat{\sigma}^2_{\widehat{m}} - \Pi_{\widehat{m}}\sigma_A^2\|_n^2 \;=\; \frac{1}{n^2}\,(R+V)^T\,\widehat{\Phi}_{\widehat{m}}\,\widehat{\Psi}_{\widehat{m}}^{-1}\,\widehat{\Phi}_{\widehat{m}}^T\,(R+V) \;\le\; \frac{1}{n^2}\,\|\widehat{\Psi}_{\widehat{m}}^{-1}\|_{\mathrm{op}}\,\Big\|\tfrac{1}{n}\,\widehat{\Phi}_{\widehat{m}}^T\,(R+V)\Big\|_{2,\widehat{m}}^2.$$

Since $\widehat{m} \in \widehat{\mathcal{M}}_{n\Delta}$, the operator norm $\|\widehat{\Psi}_{\widehat{m}}^{-1}\|_{\mathrm{op}}$ is of order $\sqrt{n\Delta}$. In addition, the map $m \mapsto \|\frac{1}{n}\,\widehat{\Phi}_m^T(R+V)\|_{2,m}^2$ is non-decreasing in m, so

$$\|\widehat{\sigma}^2_{\widehat{m}} - \Pi_{\widehat{m}}\sigma_A^2\|_n^2 \;\le\; \sqrt{\lfloor n\Delta\rfloor}\,\Big\|\tfrac{1}{n}\,\widehat{\Phi}_{\lfloor n\Delta\rfloor}^T\,(R+V)\Big\|_{2,\lfloor n\Delta\rfloor}^2.$$

By Lemma 5.4,

$$\mathbb{E}\Big[\|\widehat{\sigma}^2_{\widehat{m}} - \Pi_{\widehat{m}}\sigma_A^2\|_n^4\Big] \;\lesssim\; (n\Delta)\,\mathbb{E}\Big[\Big\|\tfrac{1}{n}\,\widehat{\Phi}_{n\Delta}^T(R+V)\Big\|_{2,n\Delta}^4\Big] \;\lesssim\; n\Delta\,(n\Delta)\,(n\Delta)^2\Big(\tfrac{1}{n^2}+\Delta^4\Big).$$

Lemma 5.5 bounds $\mathbb{P}((\Xi_{n\Delta})^c \cup (\Omega_{n\Delta})^c)$ by $(n\Delta)^{-5}$, hence

$$\mathbb{E}\Big[\|\widehat{\sigma}^2_{\widehat{m}} - \Pi_{\widehat{m}}\sigma_A^2\|_n^2\,\mathbb{1}_{(\Xi_{n\Delta})^c\cup(\Omega_{n\Delta})^c}\Big] \;\lesssim\; \frac{1}{(n\Delta)^{5/2}}\,(n\Delta)^2\,\sqrt{\tfrac{1}{n^2}+\Delta^4} \;\lesssim\; \frac{1}{n}+\Delta^2.$$

(ii) Bounding $\|\sigma_A^2\|_n^2$.
Apply Cauchy–Schwarz:

$$\mathbb{E}\Big[\|\sigma_A^2\|_n^2\,\mathbb{1}_{(\Xi_{n\Delta})^c\cup(\Omega_{n\Delta})^c}\Big] \;\le\; C\,\mathbb{E}\big[\sigma^8(X_0)\big]^{\frac12}\,\mathbb{P}\big((\Xi_{n\Delta})^c\cup(\Omega_{n\Delta})^c\big)^{\frac12} \;\lesssim\; \frac{1}{(n\Delta)^{5/2}} \;\lesssim\; \frac{1}{n}.$$

Combining the above two parts,

$$\mathbb{E}\Big[\|\widehat{\sigma}^2_{\widehat{m}} - \sigma_A^2\|_n^2\,\mathbb{1}_{(\Xi_{n\Delta}\cap\Omega_{n\Delta})^c}\Big] \;\lesssim\; \frac{1}{n}+\Delta^2.$$

2. Bound on $\mathbb{E}\big[\|\widehat{\sigma}^2_{\widehat{m}} - \sigma_A^2\|_n^2\,\mathbb{1}_{\Xi_{n\Delta}\cap\Omega_{n\Delta}}\big]$.
On $\Xi_{n\Delta}\cap\Omega_{n\Delta}$, for any $m \in \mathcal{M}_{n\Delta}$ and $t \in S_m$,

$$\widetilde{\gamma}_n(\widehat{\sigma}^2_{\widehat{m}}) + \widehat{\mathrm{pen}}(\widehat{m}) \;\le\; \widetilde{\gamma}_n(t) + \widehat{\mathrm{pen}}(m).$$

Arguing as in (19), for every $t \in S_m$,

$$\|\widehat{\sigma}^2_{\widehat{m}} - \sigma_A^2\|_n^2 \;\le\; \|t-\sigma_A^2\|_n^2 \;+\; 2\,v_n\big(\widehat{\sigma}^2_{\widehat{m}} - t\big) \;+\; 2\,\langle R,\,\widehat{\sigma}^2_{\widehat{m}} - t\rangle_n \;+\; \widehat{\mathrm{pen}}(m) \;-\; \widehat{\mathrm{pen}}(\widehat{m}).$$

Decompose $v_n(t) = \langle t, V\rangle_n$ with $V = (V_\Delta, \dots, V_{n\Delta})$ and $R = (R_\Delta, \dots, R_{n\Delta})$. Group the principal term $V^{(1)}$ and secondary terms $V^{(2)} + V^{(3)} + V^{(4)}$. Using $2\,a\,b \;\le\; \frac18\,a^2 \;+\; 8\,b^2$ one obtains

$$\|\widehat{\sigma}^2_{\widehat{m}} - \sigma_A^2\|_n^2 \;\le\; \tfrac74\,\|t-\sigma_A^2\|_n^2 \;+\; \tfrac34\,\|\widehat{\sigma}^2_{\widehat{m}} - \sigma_A^2\|_n^2 \;+\; 8\Big(\sup_{t\in B^\pi_{m,\widehat{m}}}(v_n^{(1)}(t))^2 - p(m,\widehat{m})\Big)_+$$

$$+\,8\sup_{t\in B^\pi_{m,\widehat{m}}}\overline{v_n}(t)^2 \;+\; \frac{8}{n}\sum_{k=1}^n R_{k\Delta}^2 \;+\; \widehat{\mathrm{pen}}(m) \;+\; 24\,p(m,\widehat{m}) \;-\; \widehat{\mathrm{pen}}(\widehat{m}).$$

(i) Bounding $\sup_{t\in B^\pi_{m,\widehat{m}}}(v_n^{(1)}(t))^2 - p(m,\widehat{m})$.
This term is bounded using Lemma 5.6 whose proof uses a version of Talagrand's inequality under the Berbee coupling argument to control the dependence in $(X_{k\Delta})_k$.

Lemma 5.6. *Under the assumptions of Theorem 3.1, define*

$$p(m,m') \;:=\; \kappa_0\,c_\varphi^2\,\mathbb{E}\big[\sigma^4(X_0)\big]\,\frac{(m\vee m')\,\|\Psi^{-1}_{m\vee m'}\|_{\mathrm{op}}}{n}.$$

Then there is $\kappa_0 > 0$ *such that*

$$\mathbb{E}\Big[\Big(\sup_{t\in B^\pi_{m,\hat{m}}}(v_n^{(1)}(t))^2 - p(m,\hat{m})\Big)_+\,\mathbb{1}_{\Xi_{n\Delta}\cap\Omega_{n\Delta}}\Big] \;\lesssim\; \frac{1}{n}.$$

(ii) Bounding $\sup_{t\in B^{\pi}_{m,\widehat{m}}} \overline{v_n}(t)^2$.

Since $\widehat{m}$ is bounded by $M^+_{n\Delta}$ deterministically on $\Xi_{n\Delta}$, Lemma 5.3 with $m = M^+_{n\Delta}$ implies

$$\mathbb{E}\Big[\sup_{t\in S_m+S_{\widehat{m}}} \overline{v_n}(t)^2\,\mathbb{1}_{\Xi_{n\Delta}}\Big] \;\le\; \mathbb{E}\Big[\sup_{t\in S_{M^+_{n\Delta}}} \overline{v_n}(t)^2\Big] \;\lesssim\; \Delta^2.$$

(iii) Bounding $\sum_{k=1}^n R^2_{k\Delta}/n$.

By Lemma 5.2, $\mathbb{E}[R_0^2] \lesssim \Delta^2$, so

$$\mathbb{E}\Big[\tfrac{1}{n}\sum_{k=1}^n R^2_{k\Delta}\Big] \;\lesssim\; \Delta^2.$$

(iv) Bounding the penalty terms.

$$\mathrm{pen}(m) := \kappa\, c_\varphi^2\, \mathbb{E}\big[\sigma^4(X_0)\big]\,\frac{m\,\|\Psi_m^{-1}\|_{\mathrm{op}}}{n}, \quad p(m,m') \;\le\; \mathrm{pen}(m)+\mathrm{pen}(m'),$$

provided $\kappa \ge \kappa_0$. The lemma 5.7 below ensures that one can interchange $\widehat{\mathrm{pen}}$ with pen up to a small error.

Lemma 5.7. *According to the assumptions of Theorem 3.1, there exist positive constants $c_1, c_2 > 0$ such that for all $m \in \mathcal{M}_{n\Delta}$ and $\widehat{m} \in \widehat{\mathcal{M}}_{n\Delta}$,*

$$\mathbb{E}\big[\widehat{\mathrm{pen}}(m)\,\mathbb{1}_{\Xi_{n\Delta}\cap\Omega_{n\Delta}}\big] \;\le\; c_1\,\mathrm{pen}(m) + \tfrac{c_2}{n}, \quad \mathbb{E}\Big[\big(\mathrm{pen}(\widehat{m}) - \widehat{\mathrm{pen}}(\widehat{m})\big)_+\,\mathbb{1}_{\Xi_{n\Delta}\cap\Omega_{n\Delta}}\Big] \;\le\; \tfrac{c_2}{n}.$$

Lemma 5.7 is essentially the same as Lemma 6.7 in [7], where $n\Delta$ is replaced by n in the notation and σ^2 is replaced by σ^4. The proof relies on similar concentration arguments and is omitted.

Putting all these parts together controls every term on $\Xi_{n\Delta}\cap\Omega_{n\Delta}$. Hence

$$\mathbb{E}\Big[\|\widehat{\sigma}^2_{\widehat{m}} - \sigma_A^2\|_n^2\,\mathbb{1}_{\Xi_{n\Delta}\cap\Omega_{n\Delta}}\Big] \;\lesssim\; \frac{1}{n} + \Delta^2,$$

and combining with the bound on $\mathbb{E}[\|\widehat{\sigma}^2_{\widehat{m}} - \sigma_A^2\|_n^2\,\mathbb{1}_{(\Xi_{n\Delta}\cap\Omega_{n\Delta})^c}]$ yields the final risk estimate

$$\mathbb{E}\big[\|\widehat{\sigma}^2_{\widehat{m}} - \sigma_A^2\|_n^2\big] \;\lesssim\; \frac{1}{n} + \Delta^2.$$

This completes the proof of Theorem 3.1.

5.6. **Proof of Lemma 5.6.**

Lemma 5.8 (Talagrand's Deviation Inequality)**.** *Let n be a positive integer and let $\mathcal{F}$ be a countable collection of measurable functions. Suppose $X_1,\dots,X_n$ are independent real-valued random variables and for each $f\in\mathcal{F}$ define*

$$v_n(f) \;=\; \frac{1}{n}\sum_{i=1}^n\Big(f(X_i) \;-\; \mathbb{E}[\,f(X_i)\,]\Big).$$

Assume there exist positive constants M, H, and v such that:

(1) $\sup_{f\in\mathcal{F}} \|f\|_\infty \;\le\; M,$

(2) $\mathbb{E}\Big[\sup_{f\in\mathcal{F}}\big|v_n(f)\big|\Big] \;\le\; H,$

(3) $\sup_{f\in\mathcal{F}} \frac{1}{n}\sum_{i=1}^n \mathrm{Var}\big(f(X_i)\big) \;\le\; v.$

Define $C(\alpha) := (\sqrt{1+\alpha}-1)\wedge 1$ for each $\alpha>0$, and set $b := \frac{1}{6}$. Then, for all $\alpha>0$,

$$\mathbb{E}\Big[\Big(\sup_{f\in\mathcal{F}}\big|v_n(f)\big|^2 - 2\big(1+2\alpha\big)\,H^2\Big)_+\Big] \;\le\; \frac{4}{b}\left(\frac{v}{n}\,\exp\Big(-\,b\,\alpha\,\frac{n\,H^2}{v}\Big) + \frac{49\,M^2}{6\,C(\alpha)^2\,n^2}\,\exp\Big(-\,\frac{\sqrt{2}\,b\,C(\alpha)\,\sqrt{\alpha}}{7}\,\frac{n\,H}{M}\Big)\right).$$

This result is given in [20]. To prove Lemma 5.6, we apply Lemma 5.8 together with the Berbee coupling approximation (see, for example, Chapter 5 of [23]). Throughout this proof, let

$$U_k \;:=\; W_{(k+1)\Delta} \;-\; W_{k\Delta}, \quad \chi_k \;:=\; \frac{U_k^2}{\Delta} \;-\; 1.$$

To implement the uniform bound required by Talagrand's inequality, introduce truncated versions of χ_k and $\sigma(X_{k\Delta})$. For suitably chosen integers k_n, l_n (specified later), define

$$\chi_k^c \;:=\; \big(\tfrac{u_k^2}{\Delta}-1\big)\,\mathbb{1}_{\left|\frac{u_k^2}{\Delta}-1\right|\,\le\,k_n} \;-\; \mathbb{E}\Big[\big(\tfrac{u_k^2}{\Delta}-1\big)\,\mathbb{1}_{\left|\frac{u_k^2}{\Delta}-1\right|\,\le\,k_n}\Big],$$

$$\sigma^{2,c}(x) \;:=\; \sigma^2(x)\,\mathbb{1}_{\{|\sigma^2(x)|\le l_n\}}, \qquad \chi_k^e \;:=\; \chi_k-\chi_k^c, \qquad \sigma^{2,e}(x) \;:=\; \sigma^2(x)-\sigma^{2,c}(x).$$

Set

$$v_n^{(1)}(t) \;=\; v_{n,1}^{(1)}(t) \;+\; v_{n,2}^{(1)}(t) \;+\; v_{n,3}^{(1)}(t),$$

where

$$v_{n,1}^{(1)}(t) \;:=\; \big\langle t,\; [\sigma^{2,c}(X_{k\Delta})\,\chi_k^c]_{k=1}^n\big\rangle_n, \qquad v_{n,2}^{(1)}(t) \;:=\; \big\langle t,\; [\sigma^{2,e}(X_{k\Delta})\,\chi_k^c]_{k=1}^n\big\rangle_n,$$

$$v_{n,3}^{(1)}(t) \;:=\; \big\langle t,\; [\sigma^{2}(X_{k\Delta})\,\chi_k^e]_{k=1}^n\big\rangle_n.$$

A standard application of Young's inequality shows

$$\Big(\sup_{t\in B^\pi_{m,\widehat m}}\big(v_n^{(1)}(t)\big)^2-p(m,\widehat m)\Big)_+ \;\le\; \Big(3\sup_{t\in B^\pi_{m,\widehat m}}\big(v_{n,1}^{(1)}(t)\big)^2-p(m,\widehat m)\Big)_+ + 3\sup_{t\in B^\pi_{m,\widehat m}}\big(v_{n,2}^{(1)}(t)\big)^2+3\sup_{t\in B^\pi_{m,\widehat m}}\big(v_{n,3}^{(1)}(t)\big)^2.$$

Bounding the terms $v_{n,2}^{(1)}$ and $v_{n,3}^{(1)}$ and choosing l_n and k_n.

It can be shown that

$$\mathbb{E}\Big[\sup_{t\in B^\pi_{m,\widehat m}}\big(v_{n,2}^{(1)}(t)\big)^2\,\mathbb{1}_{\Xi_n}\Big] \;\lesssim\; \frac{1}{\log^2(n\Delta)\,l_n^{25}}.$$

Indeed, with the same argument as in Lemma 5.3(ii), we can show

$$\mathbb{E}\Big[\sup_{t\in B^\pi_{m,\widehat m}}\big(v_{n,2}^{(1)}(t)\big)^2\,\mathbb{1}_{\Xi_n}\Big] \le \frac{c_\varphi^2\,M^+_{n\Delta}\,\|\Psi^{-1}_{M^+_{n\Delta}}\|_{\mathrm{op}}}{n}\,\mathbb{E}[(\chi_1^c)^2]\,\mathbb{E}[\sigma^{2,e}(X_0)],$$

then, noting that $\mathbb{E}[(\chi_1^c)^2]\le 2$ and $\|\Psi^{-1}_{M^+_{n\Delta}}\|_{\mathrm{op}}\le\frac{n\Delta}{\log^2(n\Delta)}$. Then

$$\mathbb{E}\Big[\sup_{t\in B^\pi_{m,\widehat m}}\big(v_{n,2}^{(1)}(t)\big)^2\,\mathbb{1}_{\Xi_n}\Big] \;\lesssim\; \frac{1}{\log^2(n\Delta)}\,\mathbb{E}\Big[\sigma^4(X_0)\,\mathbb{1}_{\{|\sigma^2(X_0)|\ge l_n\}}\Big].$$

Since $\mathbb{1}_{|\sigma^2(x)|\ge l_n}\le|\sigma(x)|^{50}/l_n^{25}$,

$$\mathbb{E}\Big[\sup_{t\in B^\pi_{m,\widehat m}}\big(v_{n,2}^{(1)}(t)\big)^2\,\mathbb{1}_{\Xi_n}\Big] \;\lesssim\; \frac{\mathbb{E}[|\sigma(X_0)|^{54}]}{\log^2(n\Delta)\,l_n^{25}}.$$

An analogous argument shows

$$\mathbb{E}\Big[\sup_{t\in B^\pi_{m,\widehat m}}\big(v_{n,3}^{(1)}(t)\big)^2\,\mathbb{1}_{\Xi_n}\Big] \;\lesssim\; \frac{1}{\log^2(n\Delta)\,k_n^{25}}.$$

Choose $l_n=k_n\approx\frac{(n\Delta)^{1/10}}{\log^{2/25}(n\Delta)}$ so that

$$\frac{1}{\log^2(n\Delta)\,l_n^{25}} \;\lesssim\; \frac{1}{(n\Delta)^{5/2}} \;\lesssim\; \frac{1}{n}, \quad \text{and similarly for} \quad \frac{1}{k_n^{25}}.$$

Thus,

$$\mathbb{E}\Big[\sup_{t\in B^\pi_{m,\widehat m}}\big(v_{n,2}^{(1)}(t)\big)^2\,\mathbb{1}_{\Xi_n}\Big] \;\lesssim\; \frac{1}{n}, \qquad \mathbb{E}\Big[\sup_{t\in B^\pi_{m,\widehat m}}\big(v_{n,3}^{(1)}(t)\big)^2\,\mathbb{1}_{\Xi_n}\Big] \;\lesssim\; \frac{1}{n}.$$

Bounding the term with $v_{n,1}^{(1)}(t)$.

It remains to bound $\big(3\sup_{t\in B^\pi_{m,\widehat m}}(v_{n,1}^{(1)}(t))^2-p(m,\widehat m)\big)_+$ by invoking Talagrand's inequality together with Berbee's coupling.

Berbee's coupling. Now, we employ the coupling method (Lemma 5.1 in [23]) to construct associated variables $(V^*_{i\Delta})$, where each $V_k=(\chi_k,X_{k\Delta})$ for $k=1,\dots,n$. For simplicity, assume that $n=2p_nq_n$ for some integers p_n and q_n. Then, there exist random variables $V^*_{i\Delta}=(\chi^*_k,X^*_{k\Delta})$, for $i=1,\dots,n$, satisfying the following properties:

- For each $\ell=0,\dots,p_n-1$, the random vectors

$$\vec{V}_{\ell,1} = \big(V_{(2\ell q_n+1)\Delta}, \dots, V_{(2\ell+1)q_n\Delta}\big)^T \quad \text{and} \quad \vec{V}^*_{\ell,1} = \Big(V^*_{(2\ell q_n+1)\Delta}, \dots, V^*_{(2\ell+1)q_n\Delta}\Big)^T$$

are identically distributed. Similarly, random vectors

$$\vec{V}_{\ell,2} = \big(V_{\lfloor(2\ell+1)q_n+1\rfloor\Delta}, \dots, V_{(2\ell+2)q_n\Delta}\big)^T \quad \text{and} \quad \vec{V}^*_{\ell,2} = \Big(V^*_{\lfloor(2\ell+1)q_n+1\rfloor\Delta}, \dots, V^*_{(2\ell+2)q_n\Delta}\Big)^T$$

also share the same distribution.

- For each $\ell = 0, \dots, p_n - 1$,

$$\mathbb{P}\left[\vec{V}_{\ell,1} \neq \vec{V}^*_{\ell,1}\right] \leq \beta_V\left(q_n\Delta\right) = \beta_X\left(q_n\Delta\right) \quad \text{and} \quad \mathbb{P}\left[\vec{V}_{\ell,2} \neq \vec{V}^*_{\ell,2}\right] \leq \beta_V\left(q_n\Delta\right) = \beta_X\left(q_n\Delta\right).$$

Application of Talagrand. On the event Ω^*, replace $(u_k, X_{k\Delta})$ by $(u^*_k, X^*_{k\Delta})$ and split the indices into odd and even blocks. For instance, the term

$$\mathbb{E}\Big[\Big(\sup_{t\in B^\pi_{m,\widehat{m}}} \big(v^{*,1}_{n,1}(t)\big)^2 \;-\; \tfrac{1}{6}\,p(m,m')\Big)_+\Big]$$

appears, with

$$v^{*,1}_{n,1}(t) \;:=\; \frac{1}{p_n}\sum_{\ell=0}^{p_n-1}\Bigg(\frac{1}{2\,q_n\,\Delta}\sum_{r=1}^{q_n}\chi^{c,*}_{2\,\ell\,q_n+r}\,\sigma^{2,c}\big(X^*_{(2\,\ell\,q_n+r)\Delta}\big)\,t\big(X^*_{(2\,\ell\,q_n+r)\Delta}\big)\Bigg),$$

where $\{\chi^{c,*}_k \;:=\; \chi^*_k\,\mathbb{1}_{\big|\{\chi^*_k|\,\leq\,k_n}\}$. This is a sum of p_n independent blocks. Define

$$f^{(t)}\colon (z,x) \;\mapsto\; \tfrac{1}{2\,q_n}\sum_{r=1}^{q_n}\Big[z_r\,\mathbb{1}_{|z_r|\leq k_n} \;-\; \mathbb{E}\big(z_r\,\mathbb{1}_{|z_r|\leq k_n}\big)\Big]\,t(x_r)\,\sigma^2(x_r)\,\mathbb{1}_{|\sigma^2(x_r)|\leq l_n}.$$

Hence

$$v^{*,1}_{n,1}(t) \;=\; \frac{1}{p_n}\sum_{\ell=0}^{p_n-1}\Big(f^{(t)}(Y_\ell) - \mathbb{E}[f^{(t)}(Y_\ell)]\Big),$$

where

$$Y_\ell \;=\; \big(\chi^{c,*}_{2\,\ell\,q_n+r},\, X^*_{(2\,\ell\,q_n+r)\Delta}\big)_{r=1,\dots,q_n}.$$

Apply Talagrand's inequality with $\mathcal{F} = \{f^{(t)} : t \in B^\pi_{m,m'}\}$ and $v_n(f) = v^{*,1}_{n,1}(t)$. As in Lemma 6.6 of [7], it holds that:

$$\begin{aligned}
&\sup_{f\in\mathcal{F}}\|f\|_\infty \;\leq\; M \;:=\; c_\varphi\,k_n\,l_n\,\sqrt{(m\vee m')\,\|\Psi^{-1}_{m\vee m'}\|_{\mathrm{op}}},\\
&\mathbb{E}\big[\sup_{f\in\mathcal{F}}|v_n(f)|^2\big] \;\leq\; H^2 \;:=\; \frac{2\,\mathbb{E}[\sigma^4(X_0)]\,c^2_\varphi\,(m\vee m')\,\|\Psi^{-1}_{m\vee m'}\|_{\mathrm{op}}}{n},\\
&\sup_{f\in\mathcal{F}}\frac{1}{n}\sum_{\ell=1}^{p_n}\mathrm{Var}\big(f(Y_\ell)\big) \;\leq\; v \;:=\; \frac{2\,c_\varphi}{q_n}\,\mathbb{E}[\sigma^8(X_0)]^{1/2}\,\sqrt{(m\vee m')\,\|\Psi^{-1}_{m\vee m'}\|_{\mathrm{op}}}.
\end{aligned}$$

Denote by C a universal constant. Then Talagrand's inequality (Lemma 5.8) implies

$$\mathbb{E}\Big[\big(\sup_{t\in B^\pi_{m,m'}}(v^{*,1}_{n,1}(t))^2 - 6\,H^2\big)_+\Big] \;\leq\; C_1\Big(\frac{v}{p_n}\exp\big(-C_2\,\tfrac{p_n\,H^2}{v}\big) \;+\; \frac{M^2}{p_n^2}\exp\big(-C_3\,\tfrac{p_n\,H}{M}\big)\Big).$$

Since $2\,p_n\,q_n = n$, $p_n \approx \frac{n\Delta}{\log(n\Delta)}$, and $\max\{m,m'\}\,\|\Psi^{-1}_{m\vee m'}\|_{\mathrm{op}} \lesssim \frac{n\Delta}{\log^2(n\Delta)}$, it follows that

$$\mathbb{E}\Big[\big(\sup_{t\in B^\pi_{m,m'}}(v^{*,1}_{n,1}(t))^2 - 6\,H^2\big)_+\Big] \;\leq\; C'_1\Big(\frac{\sqrt{Z}}{p_n\,q_n}\exp(-C'_2\,\sqrt{Z}) \;+\; \frac{(k_n\,l_n)^2\,Z}{p_n^2}\exp\big(-C'_3\,\tfrac{p_n}{\sqrt{n}\,k_n\,l_n}\big)\Big) := A + B,$$

where $Z := (m\vee m')\,\|\Psi^{-1}_{m\vee m'}\|_{\mathrm{op}}$. Moreover,

$$p_n\,q_n \;=\; \tfrac{n}{2}, \quad \frac{n\,\Delta}{\log(n\Delta)\,\sqrt{n}\,k_n\,l_n} \;\geq\; D_1\,n^{1/50} \;\geq\; D_2\,\log(n),$$

so the second exponential is at most $\exp(-3\log(n)) = n^{-3}$. A brief calculation then shows the series $\sum_{m'} \sqrt{Z}\exp(-c\sqrt{Z})$ converges then $A \lesssim 1/n$, while B is also $O(1/n)$. Summing over $m' \in \mathcal{M}_n^+$ yields

$$\sum_{m'} \mathbb{E}\Big[\big(\sup_{t\in B^{\pi}_{m,m'}} (v^{*,1}_{n,1}(t))^2 - \tfrac{1}{6}\,p(m,m')\big)_+\Big] \;\lesssim\; \frac{1}{n}.$$

Lastly, it suffices to control

$$\mathbb{E}\Big[\Big(\sup_{t\in B^{\pi}_{m,\widehat{m}}} v^{(1)}_n(t)^2 - p(m,\widehat{m})\Big)_+ \mathbb{1}_{\{\Xi_n\cap\Omega_n\cap(\Omega^*)^c\}}\Big],$$

which is straightforward using $\mathbb{P}((\Omega^*)^c) \lesssim 1/(n\Delta)^5$ and arguments in Lemma 5.3.

This completes the proof of Lemma 5.6.

References

1. Aurélien Alfonsi, *On the discretization schemes for the cir (and bessel squared) processes*, Monte Carlo Methods Appl. **11** (2005), no. 4, 355–384.
2. Federico M Bandi and Peter CB Phillips, *Fully nonparametric estimation of scalar diffusion models*, Econometrica **71** (2003), no. 1, 241–283.
3. Andrew Barron, Lucien Birgé, and Pascal Massart, *Risk bounds for model selection via penalization*, Probability theory and related fields **113** (1999), 301–413.
4. Fabienne Comte and Valentine Genon-Catalot, *Nonparametric drift estimation for iid paths of stochastic differential equations*, The Annals of Statistics **48** (2020), no. 6, 3336–3365.
5. ———, *Regression function estimation as a partly inverse problem*, Annals of the Institute of Statistical Mathematics **72** (2020), 1023–1054.
6. ———, *Regression function estimation on non compact support in an heteroscesdastic model*, Metrika **83** (2020), no. 1, 93–128.
7. ———, *Drift estimation on non compact support for diffusion models*, Stochastic Processes and their Applications **134** (2021), 174–207.
8. Fabienne Comte, Valentine Genon-Catalot, and Yves Rozenholc, *Penalized nonparametric mean square estimation of the coefficients of diffusion processes*, Bernoulli **13** (2007), no. 2, 514 – 543.
9. Christophe Denis, Charlotte Dion-Blanc, and Miguel Martinez, *A ridge estimator of the drift from discrete repeated observations of the solution of a stochastic differential equation*, Bernoulli **27** (2021), no. 4, 2675–2713.
10. Eddy Ella-Mintsa, *Nonparametric estimation of the diffusion coefficient from iid sde paths*, Statistical Inference for Stochastic Processes (2024), 1–56.
11. Danielle Florens, *Estimation of the diffusion coefficient from crossings*, Statistical Inference for Stochastic Processes **1** (1998), 175–195.
12. Danielle Florens-Zmirou, *On estimating the diffusion coefficient from discrete observations*, Journal of applied probability **30** (1993), no. 4, 790–804.
13. Valentine Genon-Catalot, Catherine Laredo, and Dominique Picard, *Non-parametric estimation of the diffusion coefficient by wavelets methods*, Scandinavian Journal of Statistics (1992), 317–335.
14. Arnaud Gloter, *Discrete sampling of an integrated diffusion process and parameter estimation of the diffusion coefficient*, ESAIM: Probability and Statistics **4** (2000), 205–227.
15. Marc Hoffmann, *Adaptive estimation in diffusion processes*, Stochastic processes and their Applications **79** (1999), no. 1, 135–163.
16. Marc Hofmann, *Lp estimation of the diffusion coefficient*, Bernoulli **5** (1999), no. 3, 447 – 481.
17. Jack Indritz, *An inequality for hermite polynomials*, Proceedings of the American Mathematical Society **12** (1961), no. 6, 981–983.
18. Jean Jacod, *Non-parametric kernel estimation of the coefficient of a diffusion*, Scandinavian Journal of Statistics **27** (2000), no. 1, 83–96.
19. George J Jiang and John L Knight, *A nonparametric approach to the estimation of diffusion processes, with an application to a short-term interest rate model*, Econometric Theory **13** (1997), no. 5, 615–645.
20. Thierry Klein and Emmanuel Rio, *Concentration around the mean for maxima of empirical processes*, (2005).
21. Yury A Kutoyants, *Statistical inference for ergodic diffusion processes*, Springer Science & Business Media, 2013.
22. Étienne Pardoux and Yu Veretennikov, *On the poisson equation and diffusion approximation. i*, The Annals of Probability **29** (2001), no. 3, 1061–1085.
23. Emmanuel Rio et al., *Asymptotic theory of weakly dependent random processes*, vol. 80, Springer, 2017.
24. Gabor Szegö, *Orthogonal polynomials*, (1962).
25. Sundaram Thangavelu, *Lectures on hermite and laguerre expansions*, vol. 42, Princeton University Press, 1993.