# Non-parametric estimation of the diffusion coefficient from noisy data

Emeline Schmisser

# Non-parametric estimation of the diffusion coefficient from noisy data

*Laboratoire MAP5, Université Paris Descartes.
*45 rue des Saints Pères, 75270 Paris Cedex 06, France.*
*E-mail :* emeline.schmisser@mi.parisdescartes.fr

**Abstract**

We consider a diffusion process $(X_t)_{t\geq 0}$, with drift $b(x)$ and diffusion coefficient $\sigma(x)$. At discrete times $t_k = k\delta$ for $k$ from 1 to $M$, we observe noisy data of the sample path, $Y_{k\delta} = X_{k\delta} + \varepsilon_k$. The random variables $(\varepsilon_k)$ are i.i.d, centred and independent of $(X_t)$. The process $(X_t)_{t\geq 0}$ is assumed to be strictly stationary, $\beta$-mixing and ergodic. In order to reduce the noise effect, we split data into groups of equal size $p$ and build empirical means. The group size $p$ is chosen such that $\Delta = p\delta$ is small whereas $M\delta$ is large. Then, the diffusion coefficient $\sigma^2$ is estimated in a compact set $A$ in a non-parametric way by a penalized least squares approach and the risk of the resulting adaptive estimator is bounded. We provide several examples of diffusions satisfying our assumptions and we carry out various simulations. Our simulation results illustrate the theoretical properties of our estimators.

Running title : Estimation for noisy diffusions
**Keywords**: diffusion coefficient; model selection; noisy data; non-parametric estimation; stationary distribution.
**Subject Classification:** primary 62G08, secondary 62M05

## 1 Introduction

We consider the stochastic differential equation

$$dX_t = b(X_t)dt + \sigma(X_t)dW_t, \quad X_0 = \eta \tag{1}$$

where $b, \sigma : \mathbb{R} \to \mathbb{R}$ are two unknown functions, $(W_t)$ is a Brownian motion and $\eta$ a random variable independent of $(W_t)$. It is assumed that the process $(X_t)_{t\geq 0}$ is stationary and $\beta$-mixing. At discrete times $t = 0, \delta \ldots, M\delta$, we have at our disposal noisy data of the sample path, i.e, we observe

$$Y_{k\delta} = X_{k\delta} + \varepsilon_k, \tag{2}$$

where the variables $(\varepsilon_k, k \geq 0)$ are independent, identically distributed, centred and independent of the process $(X_t)$. Our aim is to estimate the diffusion coefficient $\sigma(x)$ by a non-parametric way over a compact interval $A = [a_0, a_1]$ ($A = [0, 1]$ for instance) with the asymptotic framework $M \to \infty$, $\delta = \delta_M \to 0$ and $M\delta_M \to \infty$.

In practice, the process $(X_{k\delta})$ often cannot be observed. This may be due to either measurement devices or to what is called micro-structure noise for financial data. Several articles deal with the estimation of the diffusion coefficient for noisy data. Zhang *et al.* (2005), Rosenbaum

(2007) and Jacod *et al.* (2009) study the estimation of the integrated volatility $\int_0^1 \sigma_t^2 dt$ in a fixed time interval (here, $M\delta_M = 1$). In the same context, Gloter and Jacod (2001) study the estimation of unknown parameters in the diffusion coefficient for high frequency noisy data. Comte *et al.* (2007) study non-parametric estimation of the diffusion coefficient for non noisy data with the same asymptotic framework as ours, and Schmisser (2009b) estimates in a non-parametric way the drift of a noisy process.

In order to reduce noise effects, data is split into groups of equal size $p$ and empirical means are built as follows. Let us assume that $M = (n+2)p$, set $N = np$, $\Delta = p\delta$ and for $k = 0, \dots, n+1$,

$$\bar{Y}_{k\Delta} = \bar{X}_{k\Delta} + \bar{\varepsilon}_k$$

where

$$\bar{X}_{k\Delta} = \frac{1}{p}\sum_{j=1}^{p} X_{k\Delta+j\delta}, \qquad \bar{Y}_{k\Delta} = \frac{1}{p}\sum_{j=1}^{p} Y_{k\Delta+j\delta} \quad \text{and} \quad \bar{\varepsilon}_k = \frac{1}{p}\sum_{j=1}^{p} \varepsilon_{kp+j}. \tag{3}$$

The group size $p$ is chosen such that $\Delta = p\delta$ is small whereas $p$ is large. Then, based on the mean sample $\left(\bar{Y}_{k\Delta}, k = 0, \dots, n+1\right)$, the method of Comte *et al.* (2007) is applied to the variables $\left(\bar{Y}_{k\Delta}\right)$. First, an adequate regression-type equation is found :

$$\frac{3}{2}\frac{\left(\bar{Y}_{(k+1)\Delta} - \bar{Y}_{k\Delta}\right)^2}{\Delta} = \sigma^2\left(\bar{Y}_{(k-1)\Delta}\right) + \text{noise} + \text{remainder}.$$

(The index lag in the right hand side of the previous equation is here to avoid cumbersome correlations). Then, a penalised least-squares approach is used to build a non-parametric adaptive estimator of $\sigma^2(x)$. A family of finite-dimensional subspaces $(S_m)$ of $L^2(A)$ is introduced and a collection $\left(\hat{\sigma}_m^2\right)$ of estimators of $\sigma_A^2 = \sigma^2 \mathbb{1}_A$ is defined. Finally, after introducing a penalty, an adaptive estimator $\hat{\sigma}_{\hat{m}}^2$ is selected through a data-driven procedure among this collection.

In Section 2, the model and the assumptions are specified. In Section 3, the regression-type equation and the construction of the collection $\left(\hat{\sigma}_m^2\right)$ of estimators are precise. Theorem 1 gives the risk bound of an estimator $\hat{\sigma}_m^2$ for fixed $m$. Theorems 2 and 3 give the risk bound for the adaptive estimator $\hat{\sigma}_{\hat{m}}^2$. In Section 4, some examples of models with different noises are proposed and the estimation method is implemented on simulated data. Proofs are gathered in Section 5 and in the Appendix.

# 2 Model and assumptions

## 2.1 Assumptions

Recall that $A = [a_0, a_1]$ is the compact set where $\sigma^2$ is estimated and consider the following assumptions:

**Assumption 1.**
*Functions $\sigma(x)$ and $b(x)$ are globally Lipschitz.*

**Assumption 2.**
*There exist constants $r > 0$ and $\alpha \geq 1$ such that*

$$\exists M_0 \in \mathbb{R}^+, \quad \forall x, |x| \geq M_0, \quad xb(x) \leq -r\,|x|^\alpha .$$

**Assumption 3.**
*Diffusion coefficient $\sigma$ is bounded from below and above:*

$$\exists \sigma_1^2, \sigma_0^2 > 0, \quad \forall x \in \mathbb{R}, \quad \sigma_1^2 \leq \sigma^2(x) \leq \sigma_0^2.$$

Under Assumptions 1-3, there exists a unique invariant density $\pi$ associated to the stochastic differential equation (1).

**Assumption 4.**
*The process $(X_t)_{t\geq 0}$ is stationary:*

$$\eta \sim \pi$$

*where*

$$\pi(x) \propto \frac{1}{\sigma^2(x)} \exp\left(2\int^x \frac{b(u)}{\sigma^2(u)}du\right).$$

Moreover, $\pi$ has moments of any order.

Under Assumptions 2-3, according to Pardoux and Veretennikov (2001) (Proposition 1 p.1063), the process $(X_t)$ is exponentially $\beta$-mixing: there exist positive constants $C, \theta$ such that, for any positive $t$,

$$\beta_X(t) \leq Ce^{-\theta t},$$

where $\beta_X(t)$ is the $\beta-$mixing coefficient of $(X_t)$. Moreover, Gloter (2000), Proposition A p.226, gives the following result:

**Proposition 1.**
*Under Assumptions 1-4, for any integer $k \geq 1$, there exists a positive constant $c(k)$ such that, $\forall t > 0$, $\forall h$, $0 \leq h \leq 1$,*

$$\mathbb{E}\left(\sup_{s\in[t,t+h]} |X_s - X_t|^k\right) \quad \leq \quad c(k)h^{k/2}.$$

*According to Assumption 3, the diffusion coefficient $\sigma$ is bounded and Lipschitz, so the function $x \to \sigma^2(x)$ is Lipschitz and*

$$\mathbb{E}\left(\sup_{s\in[t,t+h]} \left|\sigma^2(X_s) - \sigma^2(X_t)\right|^k\right) \leq c(k)h^{k/2}.$$

**Assumption 5.**
*The random variables $\varepsilon_k$ have density $f$, are centred and have moments of order 8. Let us denote $\tau^2 = \mathbb{E}\left(\varepsilon_1^2\right)$.*

For $(p, \delta)$ fixed, processes $\left(\bar{X}_{k\Delta}\right)$ and $\left(\bar{Y}_{k\Delta}\right)$ defined in (3) are strictly stationary. Their invariant density are respectively denoted $\bar{\pi}_{p,\delta}$ and $\tilde{\pi}_{p,\delta}$. An additional assumption on $\tilde{\pi}_{p,\delta}$ is needed:

**Assumption 6.**
*There exists positive constants $\tilde{\pi}_0$ and $\tilde{\pi}_1$ independent of $p$ and $\delta$ such that:*

$$\forall x \in \mathbb{R}, \tilde{\pi}_{p,\delta}(x) \leq \tilde{\pi}_1 \quad \textit{and} \quad \forall x \in A, \tilde{\pi}_{p,\delta}(x) \geq \tilde{\pi}_0.$$

Sufficient conditions for Assumption 6 can be found in Schmisser (2009b) (Propositions 1 and 2).

## 2.2 Approximation spaces

Our aim is to estimate the diffusion coefficient $\sigma$ over a compact set $A = [a_0, a_1]$ of $\mathbb{R}$. For simplicity, let us set $A = [0,1]$. Below, a family of nested linear subspaces $(S_{m,r})_{m \in \mathscr{M}_n(r)}$ depending on an integer $r$ included in $L^2(A)$ and of finite dimension is constructed. The integer $\mathscr{M}_n(r)$ is the index set of the collection:

$$\mathscr{M}_n(r) := \{m,\ D_{m,r} := \dim(S_{m,r}) \leq \mathscr{D}_n(r)\}$$

where the maximal dimension $\mathscr{D}_n(r)$ will be specified later. For any $m \in \mathscr{M}_n(r)$, an estimator $\hat{\sigma}^2_m$ of $\sigma^2_A := \sigma^2 \mathbb{1}_A$ belonging to $S_{m,r}$ is computed. Then the "best" possible estimator is chosen by introducing a penalty function $pen(m,r)$.

Spline functions are used in order to construct the spaces $S_{m,r}$. The spline function of degree $r$, denoted $g_r$, is the convolution $r+1$ times of the indicator function of $[0,1]$. It is $C^{r-1}$, piecewise $C^r$, and supported on $[0, r+1]$. For a fixed $r$, the linear subspaces $(S_{m,r},\ m \geq 0)$ are generated by translating/dilating $g_r$ and restriction to $[0,1]$. We set $S_{m,r} := \text{Vect}\{(f_{r,m,k}), k = -r, \ldots, 2^m - 1\}$ where

$$f_{r,m,k}(x) = 2^{m/2} g_r(2^m x - k)\, \mathbb{1}_{[0,1]}(x).$$

**Proposition 2.**

**(i)** *Functions $(f_{r,m,k})_{k=-r,\ldots,2^m-1}$ are linearly independent. As a consequence, $D_{m,r} := \dim(S_{m,r}) = 2^m + r$.*

**(ii)** *The infinite norm and the $L^2$ norm are connected: there exists a constant $\phi_1 > 0$ such that, for any function $t \in S_{m,r}$:*

$$\|t\|^2_\infty \leq \phi_1^2 D_{m,r} \|t\|^2_{L^2}$$

*where $\|t\|^2_{L^2} = \int_A t^2(x)dx$.*

**(iii)** *There exists an orthonormal basis $\psi_\lambda$ of $S_{m,r}$ such that*

$$\forall \lambda, \quad card(\lambda',\ \|\psi_\lambda \psi_{\lambda'}\|_\infty \neq 0) \leq 3(r-1).$$

**(iv)** *For any function $t \in B^\alpha_{2,\infty}$ such that $\alpha \leq r$, there exists a constant $C > 0$ such that*

$$\|t - t_m\|_{L^2} \leq 2^{-m\alpha} C.$$

*where $t_m$ is the orthogonal projection ($L^2$) of $t\mathbb{1}_A$ over the space $S_{m,r}$.*

Points **(i)** and **(iii)** are demonstrated in Schmisser (2009b) (Propositions 3 and 4), point **(ii)** is proved in Schmisser (2009a) section 6.2. Point **(iv)** is demonstrated in Meyer (1990), Proposition 4p.50.

From now on, $r$ is fixed and omitted in the notations ($S_{m,r} = S_m$, $pen(m,r) = pen(m)$, $D_{m,r} = D_m, \ldots$)

# 3 Estimation

## 3.1 Notations

Consider the normalised increment

$$T_{k\Delta} = \frac{3}{2} \frac{\left(\bar{Y}_{(k+1)\Delta} - \bar{Y}_{k\Delta}\right)^2 - 2\hat{\tau}^2/p}{\Delta}$$

where

$$\hat{\tau}^2 = \frac{1}{2N}\sum_{k=1}^{N}\left(Y_{(k+1)\delta} - Y_{k\delta}\right)^2. \tag{4}$$

We have that

$$T_{k\Delta} = \sigma^2\left(\bar{Y}_{(k-1)\Delta}\right) + \frac{3}{2}\left(A_{k\Delta} + B_{k\Delta} + C_{k\Delta} + Z_{k\Delta} + R_{k\Delta} + S_{k\Delta}\right) \tag{5}$$

where $\sigma^2\left(\bar{Y}_{(k-1)\Delta}\right)$ is the main term, $Z_{k\Delta}$, $R_{k\Delta}$ and $S_{k\Delta}$ are centred:

$$Z_{k\Delta} = \frac{1}{\Delta}\left[\left(\int_{k\Delta+\delta}^{(k+2)\Delta} H_k^\sigma(s)dW_s\right)^2 - \int_{k\Delta+\delta}^{(k+2)\Delta}\left(H_k^\sigma(s)\right)^2 ds\right], \tag{6}$$

$$R_{k\Delta} = \frac{1}{\Delta}\left[\left(\bar{\varepsilon}_{k+1} - \bar{\varepsilon}_k\right)^2 - 2\tau^2/p\right] \tag{7}$$

$$S_{k\Delta} = 2\left(\bar{\varepsilon}_{k+1} - \bar{\varepsilon}_k\right)\frac{\bar{J}_k(\sigma)}{\Delta}. \tag{8}$$

where

$$H_k^\sigma(s) = \frac{1}{p}\sum_{j=1}^{p}\mathbb{1}_{]k\Delta+j\delta,(k+1)\Delta+j\delta]}(s)\sigma(X_s) \quad \text{and} \quad \bar{J}_k(t) = \frac{1}{p}\sum_{j=1}^{p}\int_{k\Delta+j\delta}^{(k+1)\Delta+j\delta} t(X_s)dW_s. \tag{9}$$

We have then that $\bar{J}_k(\sigma) = \int_{k\Delta+\delta}^{(k+2)\Delta} H_k^\sigma(s)dW_s$. Terms $A_{k\Delta} = A_{k\Delta}^{(1)} + A_{k\Delta}^{(2)} + A_{k\Delta}^{(3)}$, $B_{k\Delta} = B_{k\Delta}^{(1)} + B_{k\Delta}^{(2)} + B_{k\Delta}^{(3)}$ and $C_{k\Delta}$ are negligible. With

$$\bar{I}_k(t) = \frac{1}{p}\sum_{j=1}^{p}\int_{k\Delta+j\delta}^{(k+1)\Delta+j\delta} t(X_s)ds,$$

there are equal to

$$\begin{aligned} A_{k\Delta}^{(1)} &= \frac{1}{\Delta}\bar{I}_k^2(b), \qquad A_{k\Delta}^{(2)} = \frac{2}{\Delta}\bar{I}_k(b)\bar{J}_k(\sigma) \\ A_{k\Delta}^{(3)} &= \frac{1}{p^2\Delta}\sum_{i,j}\int_{k\Delta+(i\vee j)\delta}^{(k+1)\Delta+(i\wedge j)\delta}\left(\sigma^2(X_s) - \sigma^2(\bar{X}_{(k-1)\Delta})\right)ds, \end{aligned} \tag{10}$$

$$\begin{aligned} B_{k\Delta}^{(1)} &= 2\left(\bar{\varepsilon}_{k+1} - \bar{\varepsilon}_k\right)\frac{1}{\Delta}\bar{I}_k(b), \\ B_{k\Delta}^{(2)} &= \frac{2}{3}\left(\sigma^2\left(\bar{X}_{(k-1)\Delta}\right) - \sigma^2\left(\bar{Y}_{(k-1)\Delta}\right)\right) \\ B_{k\Delta}^{(3)} &= \sigma^2\left(\bar{X}_{(k-1)\Delta}\right)\left(\frac{1}{p^2\Delta}\sum_{i,j}\int_{k\Delta+(i\vee j)\delta}^{(k+1)\Delta+(i\wedge j)\delta} ds - \frac{2}{3}\right) \end{aligned} \tag{11}$$

and

$$C_{k\Delta} = 2\frac{\tau^2 - \hat{\tau}^2}{p\Delta}.$$

The index lag in equation (5) avoids cumbersome correlations ($R_{k\Delta}$ and $S_{k\Delta}$ are independent of $\bar{Y}_{(k-1)\Delta}$, not of $\bar{Y}_{k\Delta}$). Let us consider the contrast

$$\gamma_n(t) = \frac{1}{n}\sum_{k=1}^{n}\left(T_{k\Delta} - t\left(\bar{Y}_{(k-1)\Delta}\right)\right)^2$$

and define the estimator

$$\hat{\sigma}_m^2 = \arg\min_{t\in S_m} \gamma_n(t).$$

*Remark 1.* A function minimising $\gamma_n$ can always be found, but it may be not unique. The transpose of the vector $v$ is denoted $v^*$. Setting $T = (T_\Delta, \ldots, T_{n\Delta})^*$, the random vector $(\hat{b}_m(\bar{Y}_0), \ldots, \hat{b}_m(\bar{Y}_{(n-1)\Delta}))^* = \Pi_m(T)$, where $\Pi_m$ is the Euclidean projection over the subspace of $\mathbb{R}^n$ defined by $\left\{(t(\bar{Y}_0), \ldots, t(\bar{Y}_{(n-1)\Delta}),\, t \in S_m\right\}$, is always uniquely defined.

This is why, as in Comte *et al.* (2007), the risk function is chosen as the expectation of an empirical norm:

$$\mathscr{R}\left(\hat{\sigma}_m^2\right) = \mathbb{E}\left(\left\|\hat{\sigma}_m^2 - \sigma_A^2\right\|_n^2\right)$$

where $\|t\|_n^2 := \frac{1}{n}\sum_{k=1}^{n} t^2\left(\bar{Y}_{(k-1)\Delta}\right)$ is an empirical norm and $t_A := t\mathbb{1}_A$.

## 3.2 Estimation with fixed $m$

Let us specify the asymptotic framework.

**Assumption 7.**
*It is assumed that*

$$N \to \infty, \quad n \to \infty, \quad \delta \to 0, \quad p \to \infty, \quad \Delta \to 0 \quad np^2\Delta^2 \to \infty$$

*and*

$$\frac{\ln^2(n)}{n\Delta} \to 0, \qquad \mathscr{D}_n \leq \frac{cn\Delta}{\ln^2(n)}$$

*Remark 2.* Here, the last two inequalities imply that $n\Delta = N\delta \to \infty$.

**Theorem 1.**
*Under Assumptions 1-7,*

$$\mathbb{E}\left(\left\|\hat{\sigma}_m^2 - \sigma_A^2\right\|_n^2\right) \leq 3\tilde{\pi}_1\left\|\sigma_m^2 - \sigma_A^2\right\|_{L^2}^2 + C\left(\Delta + \frac{1}{p} + \frac{1}{n}\left(1 + \frac{1}{p\Delta}\right)^2 + \frac{D_m}{n}\left(\sigma_0^2 + \frac{\tau^2}{p\Delta}\right)^2\right)$$

*where $\sigma_m^2$ is the orthogonal projection ($L^2$) of $\sigma^2\mathbb{1}_A$ over $S_m$.*

**Corollary 1.**
*The previous inequality reaches its minimum for $p = \Delta^{-1} = \delta^{-1/2}$. In that case, the estimator risk satisfies:*

$$\mathbb{E}\left(\left\|\hat{\sigma}_m^2 - \sigma_A^2\right\|_n^2\right) \leq 3\tilde{\pi}_1\left\|\sigma_m^2 - \sigma_A^2\right\|_{L^2}^2 + C\left(\Delta + \frac{1}{n} + \left(\sigma_0^2 + \tau^2\right)^2\frac{D_m}{n}\right).$$

## 3.3 Convergence rate

Assume that $\sigma_A^2$ belongs to a Besov space $B_{2,\infty}^{\alpha}$ and that $\left\|\sigma_A^2\right\|_{B_{2,\infty}}^2 \leq 1$. According to Proposition 2 **(iv)**, if $\alpha \leq r$, the bias term is bounded by

$$\left\|\sigma_m^2 - \sigma_A^2\right\|_{L^2}^2 \leq C 2^{-2m\alpha}.$$

Then, the bias-variance compromise $\left\|\sigma_m^2 - \sigma_A^2\right\|_{L^2}^2 + \frac{D_m}{n}$ is minimum when $m = \frac{\log_2(n)}{1+2\alpha}$, and the risk satisfies:

$$\mathbb{E}\left(\left\|\hat{\sigma}_m^2 - \sigma_A^2\right\|_n^2\right) \leq C n^{-2\alpha/(1+2\alpha)} + C\Delta.$$

*Remark 3.* Let us set $\delta \sim N^{-\beta}$. As $N\delta \to \infty$, it is needed that $0 < \beta < 1$. We have the following convergence rates:

| $\beta$ | main term | convergence speed |
|---|---|---|
| $0 < \beta \leq \frac{4\alpha}{(4\alpha+1)}$ | $\Delta$ | $N^{-\beta/2}$ |
| $\frac{4\alpha}{(4\alpha+1)} \leq \beta < 1$ | $n^{-2\alpha/(2\alpha+1)}$ | $N^{-(1-\beta/2)2\alpha/(2\alpha+1)}$ |

The convergence rate of the risk is always larger than $N^{-1/2}$. It is maximal if $\delta \sim N^{-4\alpha/(4\alpha+1)}$.

## 3.4 Adaptive estimation

Our aim is to choose $\hat{m}$ in order to minimise the bias-variance compromise. A penalty function $pen(m)$ depending on $n$ and $D_m$ is introduced, $\hat{m}$ is chosen such that

$$\hat{m} = \arg\min_{m \in \mathscr{M}_n} \left\{\gamma_n\left(\hat{\sigma}_m^2\right) + pen(m)\right\},$$

and the resulting estimator is denoted by $\hat{\sigma}_{\hat{m}}^2$ .

### 3.4.1 Additional assumption on the noise

To construct an adaptive estimator, an additional assumption on the law of $\varepsilon$ is needed:

**Assumption 8.**
*Assume that there exists some constant $\upsilon$ such that, for any $l \geq 0$:*

$$\mathbb{E}\left(\varepsilon_1^{2l}\right) \leq l^l \upsilon^{2l} \quad \textit{and} \quad \mathbb{E}\left(\varepsilon_1^{2l+1}\right) = 0$$

*Remark 4.* If variables $(\varepsilon_k)$ satisfy Assumption 8, then there are sub-Gaussian, *i.e.* for any $\lambda$, $\mathbb{E}\left(e^{\lambda\varepsilon_k}\right) \leq \exp\left(\upsilon^2\lambda^2\right)$ . Hence, one can apply Theorem 2 of (Schmisser, 2009b) which gives a bound of the risk of the drift estimator. The Laplace law does not satisfy Assumption 8, contrary to Gaussian laws and bounded laws.

**Theorem 2.**
*Under Assumptions 1-8, there exists a constant $\kappa$ such that, setting*

$$pen(m) \geq \frac{\kappa D_{m,r}}{n}\left(\sigma_0^2 + \frac{\upsilon^2}{p\Delta}\right)^2$$

*we have that, for $p \sim \delta^{-1/2}$:*

$$\mathbb{E}\left(\left\|\hat{\sigma}_{\hat{m}}^2 - \sigma_A^2\right\|_n^2\right) \leq C \inf_{m \in \mathscr{M}_n}\left(\left\|\hat{\sigma}_m^2 - \sigma_A^2\right\|_{L^2}^2 + pen(m)\right) + \frac{C}{n} + C'\delta^{1/2}.$$

Parameters $\sigma_0^2$ and $\upsilon^2$ are unknown, but they can be replaced by rough estimators.

### 3.4.2 Without Assumption (8)

If variables $(\varepsilon_k)$ do not satisfy Assumption 8, the inequality obtained in Theorem 1 can not be generalised to the adaptive case. However, we can obtain a weaker result.

**Theorem 3.**
*Under Assumptions 1-7, if $p \sim \delta^{-2/3}$, there exists a constant $\kappa$ such that, if*

$$pen(m) \geq \kappa \sigma_0^4 \frac{D_{m,r}}{n},$$

*then*

$$\mathbb{E}\left(\left\|\hat{\sigma}^2_{\hat{m}} - \sigma^2_A\right\|_n^2\right) \leq C \inf_{m \in \mathscr{M}_n} \left(\left\|\sigma^2_m - \sigma^2_A\right\|^2_{L^2} + pen(m)\right) + \frac{C}{n} + C'\delta^{1/3}.$$

# 4 Simulations

**Simulation algorithm**

1. Simulate a trajectory $\left(\bar{X}_{k\Delta}\right)$ using a Euler scheme of step $\delta$. Figures 1 and 2 are drawn for $n = 10^4$, $\Delta = 10^{-2}$ and $p = 100$ (so $N = 10^6$ and $\delta = 10^{-4}$). This may seem very large, but Jacod *et al.* (2009), Rosenbaum (2007) and Zhang *et al.* (2005) use approximately 20000 data to estimate the integrated volatility with a microstructure noise (one data every second for one day). Here, we estimate $\sigma^2$ in a non-parametric way, so we need more data.

2. Simulate the random variables $(\bar{\varepsilon}_k)$.

3. Construct $\bar{Y}_{k\Delta} = \bar{X}_{k\Delta} + \bar{\varepsilon}_k$ and

$$T_{k\Delta} = \frac{\left(\bar{Y}_{(k+1)\Delta} - \bar{Y}_{k\Delta}\right)^2 - 2\tau^2}{\Delta}$$

   (as the noise variance is known, it is not necessary to estimate it).

4. To plot the figures, eliminate the 5% extreme values.

5. For any $m \in \mathscr{M}_n$, compute $\hat{\sigma}_{m,1} \in S_{m,1}$ and $\hat{\sigma}_{m,2} \in S_{m,2}$.

6. Compute the penalty function

$$pen(m,r) = \frac{\kappa D_{m,r}}{n}\left(\sigma_0^4 + \frac{\upsilon^4}{p^2\Delta^2}\right)$$

   The parameter $\upsilon$ is defined in Assumption 8. It is chosen such that $\upsilon^2 > 2\hat{\tau}^2$. In fact, for Gaussian variables of law $\mathscr{N}(0,\tau^2)$, the parameter $\upsilon^2$ is such that $\upsilon^2 = 2\tau^2$. The parameter $\kappa$ is chosen equal to 13 by numerical calibration (see Comte and Rozenholc (2002) for more explanations). The parameter $\sigma_0^2$ is replaced with a rough estimator. We don't want to under-estimate $\sigma_0^2$, so we replace it by something much larger, for instance $2\hat{\sigma}^2_{0,0} + 2\sup_{x\in A}\hat{\sigma}^2_{0,1}(x)$ (the estimator $\hat{\sigma}^2_{0,0}$ is a constant, whereas $\hat{\sigma}^2_{0,1}$ is a linear function).

7. Compute the estimators over $(S_{m,1},\ m \geq 0)$ and $(S_{m,2},\ m \geq 0)$ respectively denoted by $\hat{\sigma}_{\hat{m},1}$ and $\hat{\sigma}_{\hat{m},2}$. The two estimators are plotted on the figures, but only $\hat{\sigma}_{\hat{m},1}$ is used in the tables.

The following table presents, for each model, the drift and the diffusion coefficient functions, the estimation interval and some remarks.

| Model | $b(x)$ | $\sigma(x)$ | Estimation interval | Remarks |
|---|---|---|---|---|
| Model 1 | $-2x$ | $1$ | $[-1,1]$ | Ornstein Uhlenbeck $\sigma \in S_0$ |
| Model 2 | $-2x$ | $\sqrt{\frac{1+x^2}{2+x^2}}$ | $[-1,1]$ | |
| Model 3 | $-2x$ | $\frac{1}{\sqrt{1+x^2}}$ | $[-1,1]$ | $\sigma$ not bounded from below |
| Model 4 | $-2x$ | $1+\lvert x\rvert$ | $[-2,2]$ | $\sigma$ not bounded from above |

We use three different noises, which are presented here.

| law | $\tau^2$ | $\upsilon^2$ | satisfies Assumption 8 |
|---|---|---|---|
| Gaussian $\mathcal{N}(0,1)$ | 1 | 2 | yes |
| Uniform over $[-1,1]$ | $1/3$ | 1 | yes |
| Laplace: $f(x) = 1/2\exp(-\lvert x\rvert)$ | 2 | 5 (arbitrary) | no |

In order to construct the tables, 50 trajectories are simulated for each noise and each model. The mean of the error between the estimated and the true diffusion coefficient is estimated by the empirical norm

$$ris = \text{mean}\left(\left\|\hat{\sigma}_{\hat{m},r} - \sigma_A^2\right\|_n^2\right).$$

Moreover, in order to check whether the algorithm is adaptive, an oracle is computed:

$$oracle = \text{mean}\left(\frac{\left\|\hat{\sigma}_{\hat{m},r} - \sigma_A^2\right\|_n^2}{\min_m\left(\left\|\hat{\sigma}_{m,r} - \sigma_A^2\right\|_n^2\right)}\right).$$

The minima of the empirical risk $ris$ and of the oracle $or$ are in bold.

**Results:** The greater $n$ is, the better the estimators. In particular, Tables 1-4 show that the smallest empirical risks are obtained for $n = 10^4$ and $\Delta = 10^{-2}$ or $\Delta = 3.10^{-3}$ (so $N$ larger than $10^6$). The oracle is better for small values of $n$: actually, the penalty function is larger so it is easier to choose the appropriate estimator. The empirical risks are smaller for the uniform noise and larger for the noise with Laplace law. This is partly because the parameter $\upsilon$ is smaller for the uniform noise (=1) and larger for the noise with Laplace law ($\simeq 5$). The Laplace noise does not satisfy Assumption 5, nevertheless, the empirical risk for this noise is comparable with the others. The estimation results are very dependent of the choice of $\upsilon^2$ and $\kappa$. So, it is better to add negligible correcting terms (see Comte and Rozenholc (2004)). Such terms avoid under-penalisation and are in accordance with the fact that the theorems provide lower bounds for the penalty. The correcting terms are negligible so they do not affect the rate of convergence.

# 5 Proofs

## 5.1 Preliminary results

In the proofs, $r$ is fixed and we denote $S_{m,r} = S_m$ and $D_{m,r} = D_m$. We have

$$\gamma_n(t) - \gamma_n(\sigma^2) = \|t - \sigma\|_n^2 - \frac{2}{n}\sum_{k=1}^{n}\left(T_{k\Delta} - \sigma^2\left(\bar{Y}_{(k-1)\Delta}\right)\right)\left(t\left(\bar{Y}_{(k-1)\Delta} - \sigma^2\left(\bar{Y}_{(k-1)\Delta}\right)\right)\right).$$

Let us recall that $\sigma_m^2$ denotes the orthogonal projection ($L^2$) of $\sigma_A^2$ over $S_m$. As $\gamma_n(\hat{\sigma}_m^2) - \gamma_n(\sigma^2) \leq \gamma_n(\sigma_m^2) - \gamma_n(\sigma^2)$, we have:

$$\left\|\hat{\sigma}_m^2 - \sigma^2\right\|_n^2 \leq \left\|\sigma_m^2 - \sigma^2\right\|_n^2 + \frac{1}{n}\sum_{k=1}^{n}\left(T_{k\Delta} - \sigma^2\left(\bar{Y}_{(k-1)\Delta}\right)\right)\left(\hat{\sigma}_m^2\left(\bar{Y}_{(k-1)\Delta}\right) - \sigma_m^2\left(\bar{Y}_{(k-1)\Delta}\right)\right).$$

Let us set

$$\nu_n(t) = \frac{1}{n}\sum_{k=1}^{n} Z_{k\Delta}\, t(\bar{Y}_{(k-1)\Delta}), \quad \rho_n(t) = \frac{1}{n}\sum_{k=1}^{n} R_{k\Delta}\, t(\bar{Y}_{(k-1)\Delta}), \tag{12}$$

$$\xi_n(t) = \frac{1}{n}\sum_{k=1}^{n} S_{k\Delta}\, t(\bar{Y}_{(k-1)\Delta}), \quad E_n(t) = \frac{1}{n}\sum_{k=1}^{n} \left(A_{k\Delta} + B_{k\Delta} + C_{k\Delta}\right) t(\bar{Y}_{(k-1)\Delta}). \tag{13}$$

(see 6-11). We can write:

$$\left\|\hat{\sigma}_m^2 - \sigma^2\right\|_n^2 \leq \left\|\sigma_m^2 - \sigma^2\right\|_n^2 + 2\left(\nu_n + \rho_n + \xi_n + E_n\right)\left(\hat{\sigma}_m^2 - \sigma_m^2\right),$$

and, as $\sigma_m^2$ and $\hat{\sigma}_m^2$ are $A$-supported:

$$\left\|\hat{\sigma}_m^2 - \sigma_A^2\right\|_n^2 \leq \left\|\sigma_m^2 - \sigma_A^2\right\|_n^2 + 2\left(\nu_n + \rho_n + \xi_n + E_n\right)\left(\hat{\sigma}_m^2 - \sigma_m^2\right). \tag{14}$$

**Lemma 1.**
*Consider the following filtrations:*

$$\mathscr{F}_t = \sigma\left(\eta,\, W_s, s \leq t\right) \quad \textit{and} \quad \mathscr{G}_t = \sigma\left(\eta,\, W_s, s \leq t,\, \varepsilon_j, j\delta \leq t\right),\ t \geq 0. \tag{15}$$

*Under Assumptions 1-5, we have:*

**(i)** *The random variables $Z_{k\Delta}^2$ are $\mathscr{G}_{(k+2)\Delta}$-measurable and $\mathbb{E}\left(Z_{k\Delta}\middle|\mathscr{G}_{k\Delta}\right) = 0$. Moreover,*

$$\mathbb{E}\left(Z_{k\Delta}^2\middle|\mathscr{G}_{k\Delta}\right) \leq c\sigma_0^4 \quad \textit{and} \quad \mathbb{E}\left(Z_{k\Delta}^4\right) \leq c\sigma_0^8.$$

**(ii)** *The random variables $R_{k\Delta}$ are $\mathscr{G}_{(k+2)\Delta}$-measurable. Moreover,*

$$\mathbb{E}\left(R_{k\Delta}^2\right) \leq \frac{c\tau^4}{p^2\Delta^2} + \frac{c\tau_4}{p^3\Delta^2} \quad \textit{and} \quad \mathbb{E}\left(R_{k\Delta}^4\right) \leq \frac{c\tau^8}{p^4\Delta^4} + \frac{c\tau_8}{p^7\Delta^4}.$$

**(iii)** *The random variables $\bar{J}_k(\sigma)$ are $\mathscr{G}_{(k+2)\Delta}$-measurable, and*

$$\mathbb{E}\left(\bar{J}_k(\sigma)\middle|\mathscr{G}_{k\Delta}\right) = 0 \quad \textit{and} \quad \mathbb{E}\left[\bar{J}_k^2(\sigma)\middle|\mathscr{G}_{k\Delta}\right] \leq c\Delta\sigma_0^2.$$

*Besides, the random variables $S_{k\Delta}$ are $\mathscr{G}_{(k+2)\Delta}$-measurable and*

$$\mathbb{E}\left(S_{k\Delta}^2\right) \leq c\tau^2\sigma_0^2/(p\Delta) \quad \textit{and} \quad \mathbb{E}\left(S_{k\Delta}^4\right) \leq c\tau^2\sigma_0^4/(p^2\Delta^2).$$

**(iv)** $\mathbb{E}\left(A_{k\Delta}^2\right) \leq c\Delta$ *and* $\mathbb{E}\left(A_{k\Delta}^4\right) \leq c\Delta^2$

**(v)** $\mathbb{E}\left(B_{k\Delta}^2\right) \leq c/p$ *and* $\mathbb{E}\left(B_{k\Delta}^4\right) \leq c/p^2$

*where* $\tau_4 = \mathbb{E}\left(\varepsilon_1^4\right)$ *and* $\tau_8 = \mathbb{E}\left(\varepsilon_1^8\right)$.
*(see* (6)-(11)*).*

The term $C_{k\Delta} = 2\frac{\hat{\tau}^2-\tau^2}{p\Delta}$ is studied thanks to the following lemma:

**Lemma 2.**
*We have*

**(i)** $\left|\mathbb{E}\left(\hat{\tau}^2-\tau^2\right)\right| \leq \delta^2\mathbb{E}\left[b^2(X_0)\right] + \delta\mathbb{E}\left(\sigma^2(X_0)\right)$.

**(ii)** $\mathbb{E}\left(\left(\hat{\tau}^2-\tau^2\right)^2\right) \leq (\tau^4+\tau_4)/N + c\delta^2$

**(iii)** $\mathbb{E}\left(\left(\hat{\tau}^2-\tau^2\right)^4\right) \leq c/N^2 + c\delta^4$.

If the noise law is known, $\hat{\tau}^2$ is replaced by $\tau^2$ and $C_{k\Delta} = 0$. Thanks to the Cauchy-Schwartz inequality, we obtain (see (13) and (14)):

$$\begin{aligned} E_n\left(\hat{\sigma}_m^2-\sigma_m^2\right) &\leq \frac{1}{12}\left\|\hat{\sigma}_m^2-\sigma_m^2\right\|_n^2 + \frac{12}{n}\sum_{k=1}^n\left(A_{k\Delta}+B_{k\Delta}+C_{k\Delta}\right)^2 \\ &\leq \frac{1}{12}\left\|\hat{\sigma}_m^2-\sigma_m^2\right\|_n^2 + \frac{36}{n}\sum_{k=1}^n\left(A_{k\Delta}^2+B_{k\Delta}^2+C_{k\Delta}^2\right). \end{aligned}$$

Let us consider the norm $\|t\|_{\tilde{\pi}}^2 = \int_A t^2(x)\tilde{\pi}(x)dx$ where $\tilde{\pi} = \tilde{\pi}_{p,\delta}$ is the density of $\bar{Y}_\Delta$. As the random variables $\bar{Y}_{k\Delta}$ have density $\tilde{\pi}$, we have $\mathbb{E}\left(\|t\|_n^2\right) = \|t\|_{\tilde{\pi}}^2$. Let us set $\mathscr{B}_m = \left\{t \in S_m, \|t\|_{\tilde{\pi}}^2 \leq 1\right\}$. We have

$$\begin{aligned} \left(\nu_n+\rho_n+\xi_n\right)\left(\hat{\sigma}_m^2-\sigma_m^2\right) &\leq \left\|\hat{\sigma}_m^2-\sigma_m^2\right\|_{\tilde{\pi}} \sup_{t\in\mathscr{B}_m}\left(\nu_n(t)+\rho_n(t)+\xi_n(t)\right) \\ &\leq \frac{1}{12}\left\|\hat{\sigma}_m^2-\sigma_m^2\right\|_{\tilde{\pi}}^2 + 36\sup_{t\in\mathscr{B}_m}\left(\nu_n^2(t)+\rho_n^2(t)+\xi_n^2(t)\right). \end{aligned}$$

Let us introduce the set

$$\Omega_n = \left\{\omega, \forall t \in \bigcup_{m,m'} S_m + S_{m'}, \quad \left|\frac{\|t\|_n^2}{\|t\|_{\tilde{\pi}}^2} - 1\right| \leq \frac{1}{2}\right\}$$

in which norms $\|.\|_n$ and $\|.\|_{\tilde{\pi}}$ are equivalent: on $\Omega_n$, we have

$$\|t\|_{\tilde{\pi}}^2 \leq 2\|t\|_n^2 \leq 3\|t\|_{\tilde{\pi}}^2.$$

## 5.2 Proof of Theorem 1

First, we study the risk on $\Omega_n$. As, on $\Omega_n$, $\|t\|_{\tilde{\pi}}^2 \leq 2\|t\|_n^2$ and $\left\|\hat{\sigma}_m^2-\sigma_m^2\right\|_n^2 \leq 2\left\|\hat{\sigma}_m^2-\sigma_A^2\right\|_n^2 + 2\left\|\sigma_m^2-\sigma_A^2\right\|_n^2$, we obtain:

$$\left\|\hat{\sigma}_m^2-\sigma_A^2\right\|_n^2 \leq 3\left\|\sigma_m^2-\sigma_A^2\right\|_n^2 + 72\sup_{t\in\mathscr{B}_m}\left(\nu_n^2(t)+\rho_n^2(t)+\xi_n^2(t)\right) + \frac{72}{n}\sum_{k=1}^n\left(A_{k\Delta}^2+B_{k\Delta}^2+C_{k\Delta}^2\right).$$

We have: $\mathbb{E}\left(\left\|\sigma_m^2-\sigma_A^2\right\|_n^2\right)=\left\|\sigma_m^2-\sigma_A^2\right\|_{\tilde{\pi}}^2\leq\tilde{\pi}_1\left\|\sigma_m^2-\sigma_A^2\right\|_{L^2}^2$. According to Lemma 1 **(iv)** and **(v)** and Lemma 2 **(ii)**,

$$\mathbb{E}\left(\left\|\hat{\sigma}_m^2-\sigma_A^2\right\|_n^2\mathbb{1}_{\Omega_n}\right)\leq 3\tilde{\pi}_1\left\|\sigma_m^2-\sigma_A^2\right\|_{L^2}^2+72\,\mathbb{E}\left(\sup_{t\in\mathscr{B}_m}\nu_n^2(t)+\rho_n^2(t)+\xi_n^2(t)\right)+c\Delta+\frac{c}{p}+\frac{c}{Np\Delta}$$

with $c\in\mathbb{R}^+$.

It remains to bound $\mathbb{E}\left(\sup_{t\in\mathscr{B}_m}\nu_n^2(t)\right)$, $\mathbb{E}\left(\sup_{t\in\mathscr{B}_m}\rho_n^2(t)\right)$ and $\mathbb{E}\left(\sup_{t\in\mathscr{B}_m}\xi_n^2(t)\right)$. Let us consider an orthonormal (for the $L^2_{\tilde{\pi}}$-norm) basis $(\varphi_\lambda)_{\lambda\in\Lambda_m}$ of $S_m$. Any function $t\in S_m$ can be written $t=\sum_{\lambda\in\Lambda_m}a_\lambda\varphi_\lambda$. Moreover, $t\in\mathscr{B}_m\Leftrightarrow\sum_{\lambda\in\Lambda_m}a_\lambda^2\leq 1$. As the function $t\to\nu_n(t)$ is linear, we have

$$\nu_n^2(t)=\left(\sum_{\lambda\in\Lambda_m}a_\lambda\nu_n(\varphi_\lambda)\right)^2\leq\left(\sum_{\lambda\in\Lambda_m}a_\lambda^2\right)\left(\sum_{\lambda\in\Lambda_m}\nu_n^2\left(\varphi_\lambda\right)\right).$$

So

$$\mathbb{E}\left(\sup_{t\in\mathscr{B}_m}\nu_n^2(t)\right)\leq\sum_{\lambda\in\Lambda_m}\mathbb{E}\left(\nu_n^2\left(\varphi_\lambda\right)\right). \tag{16}$$

One prove analogously that $\mathbb{E}\left(\sup_{t\in\mathscr{B}_m}\rho_n^2(t)\right)\leq\sum_{\lambda\in\Lambda_m}\mathbb{E}\left(\rho_n^2\left(\varphi_\lambda\right)\right)$ and $\mathbb{E}\left(\sup_{t\in\mathscr{B}_m}\xi_n^2(t)\right)\leq\sum_{\lambda\in\Lambda_m}\mathbb{E}\left(\xi_n^2\left(\varphi_\lambda\right)\right)$. According to (12), we have

$$\mathbb{E}\left(\nu_n^2\left(\varphi_\lambda\right)\right)\leq 2\mathbb{E}\left[\left(\frac{1}{n}\sum_{k\text{ even}}\varphi_\lambda(\bar{Y}_{(k-1)\Delta})Z_{k\Delta}\right)^2\right]+2\mathbb{E}\left[\left(\frac{1}{n}\sum_{k\text{ odd}}\varphi_\lambda(\bar{Y}_{(k-1)\Delta})Z_{k\Delta}\right)^2\right]$$

where the random variables $Z_{k\Delta}$ are defined in (6). The splitting into odd and even indices has the following advantage: each sum is composed of uncorrelated variables with identical distributions. We can write:

$$\begin{aligned}\mathbb{E}\left[\left(\frac{1}{n}\sum_{k\text{ even}}\varphi_\lambda(\bar{Y}_{(k-1)\Delta})Z_{k\Delta}\right)^2\right] &= \frac{1}{n^2}\sum_{k\text{ even}}\mathbb{E}\left(\varphi_\lambda^2(\bar{Y}_{(k-1)\Delta})Z_{k\Delta}^2\right)\\ &+ \frac{2}{n}\sum_{k<l\text{ even}}\mathbb{E}\left[\varphi(\bar{Y}_{(k-1)\Delta})\varphi(\bar{Y}_{(l-1)\Delta})Z_{k\Delta}\mathbb{E}\left(Z_{l\Delta}\middle|\mathscr{G}_{l\Delta}\right)\right].\end{aligned}$$

The second term is null, and, thanks to Lemma 1 **(ii)**, we have:

$$\begin{aligned}\mathbb{E}\left[\left(\frac{1}{n}\sum_{k\text{ even}}\varphi_\lambda(\bar{Y}_{(k-1)\Delta})Z_{k\Delta}\right)^2\right] &= \frac{1}{n^2}\sum_{k\text{ even}}\mathbb{E}\left[\varphi_\lambda^2(\bar{Y}_{(k-1)\Delta})\mathbb{E}\left(Z_{k\Delta}^2\middle|\mathscr{G}_{k\Delta}\right)\right]\\ &\leq \frac{c\sigma_0^4}{n}\mathbb{E}\left(\left\|\varphi_\lambda\right\|_n^2\right).\end{aligned}$$

As the functions $(\varphi_\lambda)_{\lambda\in\Lambda_m}$ are orthonormal (for the $L^2_{\tilde{\pi}}$ norm), for any $\lambda\in\Lambda_m$, $\mathbb{E}\left(\left\|\varphi_\lambda\right\|_n^2\right)=\left\|\varphi_\lambda\right\|_{\tilde{\pi}}^2=1$. It follows that:

$$\mathbb{E}\left[\left(\frac{1}{n}\sum_{k\text{ even}}Z_{k\Delta}\varphi_\lambda(\bar{Y}_{(k-1)\Delta})\right)^2\right]\leq\frac{c\sigma_0^4}{n}.$$

A similar result is obtained for $\mathbb{E}\left[\left(\frac{1}{n}\sum_{k \text{ odd}} Z_{k\Delta}\varphi_\lambda(\bar{Y}_{(k-1)\Delta})\right)^2\right]$. Hence

$$\mathbb{E}\left(\nu_n^2\left(\varphi_\lambda\right)\right) \leq \frac{c\sigma_0^4}{n},$$

and, according to (16),

$$\mathbb{E}\left(\sup_{t\in\mathscr{B}_m} \nu_n^2(t)\right) \leq \frac{c\sigma_0^4 D_m}{n}. \tag{17}$$

In the same way:

$$\mathbb{E}\left(\rho_n^2(\varphi_\lambda)\right) \leq 2\mathbb{E}\left[\left(\frac{1}{n}\sum_{k \text{ even}} R_{k\Delta}\varphi_\lambda(\bar{Y}_{(k-1)\Delta})\right)^2\right] + 2\mathbb{E}\left[\left(\frac{1}{n}\sum_{k \text{ odd}} R_{k\Delta}\varphi_\lambda(\bar{Y}_{(k-1)\Delta})\right)^2\right]$$

where the random variables $R_{k\Delta}$ are defined in (7). Let us consider $k \leq l-2$. Thanks to the lag, $R_{l\Delta}$ is independent of $R_{k\Delta}$, $\bar{Y}_{(k-1)\Delta}$, $\bar{Y}_{(l-1)\Delta}$ and $\bar{Y}_{(k-1)\Delta}$ is independent of $R_{k\Delta}$. Hence:

$$\begin{aligned}
\mathbb{E}\left[\left(\frac{1}{n}\sum_{k \text{ even}} R_{k\Delta}\varphi_\lambda(\bar{Y}_{(k-1)\Delta})\right)^2\right] &= \frac{2}{n^2}\sum_{k<l \text{ even}} \mathbb{E}(\varphi_\lambda(\bar{Y}_{(k-1)\Delta})\varphi_\lambda(\bar{Y}_{(l-1)\Delta})R_{k\Delta})\mathbb{E}(R_{l\Delta}) \\
&+ \frac{1}{n^2}\sum_{k \text{ even}} \mathbb{E}(\varphi_\lambda^2(\bar{Y}_{(k-1)\Delta}))\mathbb{E}(R_{k\Delta}^2) \\
&= \frac{1}{n^2}\sum_{k \text{ even}} \mathbb{E}(\varphi_\lambda^2(\bar{Y}_{(k-1)\Delta}))\mathbb{E}(R_{k\Delta}^2)
\end{aligned}$$

We know that $\mathbb{E}\left(\varphi_\lambda^2(\bar{Y}_{(k-1)\Delta})\right) = 1$. According to Lemma 1 **(i)**, $\mathbb{E}\left(R_{k\Delta}^2\right) \leq \frac{c\tau}{p^2\Delta^2} + \frac{c\tau_4^4}{p^3\Delta^2}$. So

$$\mathbb{E}\left(\rho_n^2(\varphi_\lambda)\right) \leq \frac{c\tau^4}{np^2\Delta^2} + \frac{c\tau_4^4}{np^3\Delta^2},$$

and

$$\mathbb{E}\left(\sup_{t\in\mathscr{B}_m} \rho_n^2(t)\right) \leq c\frac{\tau^4 D_m}{np^2\Delta^2}.$$

Similarly, we have

$$\mathbb{E}\left(\xi_n^2(\varphi_\lambda)\right) \leq 2\mathbb{E}\left[\left(\frac{1}{n}\sum_{k \text{ even}} S_{k\Delta}\varphi_\lambda(\bar{Y}_{(k-1)\Delta})\right)^2\right] + 2\mathbb{E}\left[\left(\frac{1}{n}\sum_{k \text{ odd}} S_{k\Delta}\varphi_\lambda(\bar{Y}_{(k-1)\Delta})\right)^2\right]$$

The term $S_{k\Delta} = 2\frac{\bar{\varepsilon}_{k+1}-\bar{\varepsilon}_k}{\Delta}\bar{J}_k\left(\sigma\right)$ is the product of two independent terms. Moreover, for $k \leq l+2$, $(\bar{\varepsilon}_{l+1} - \bar{\varepsilon}_l)$ is independent of $\bar{Y}_{(l-1)\Delta}$, $\bar{Y}_{(k-1)\Delta}$ and $S_{k\Delta}$. We have

$$\begin{aligned}
\mathbb{E}\left[\left(\frac{1}{n}\sum_{k \text{ even}} S_{k\Delta}\varphi_\lambda(\bar{Y}_{(k-1)\Delta})\right)^2\right] &= \frac{1}{n^2\Delta^2}\sum_{k \text{ even}} \mathbb{E}\left[\varphi_\lambda^2(\bar{Y}_{(k-1)\Delta})\mathbb{E}\left(\bar{J}_k^2(\sigma)\middle|\mathscr{G}_{k\Delta}\right)\right]\mathbb{E}\left[\left(\bar{\varepsilon}_{l+1}-\bar{\varepsilon}_l\right)^2\right] \\
&\leq \frac{c\tau^2\sigma_0^2}{np\Delta}.
\end{aligned}$$

So

$$\mathbb{E}\left(\sup_{t\in\mathscr{B}_m}\xi_n^2(t)\right)\leq c\frac{\tau^2\sigma_0^2D_m}{np\Delta}.$$

Hence:

$$\mathbb{E}\left(\left\|\hat{\sigma}_m^2-\sigma_A^2\right\|_n^2\mathbb{1}_{\Omega_n}\right)\leq 3\tilde{\pi}_1\left\|\sigma_m^2-\sigma_A^2\right\|_{L^2}^2+\frac{cD_m}{n}\left(\sigma_0^2+\frac{\tau^2}{p\Delta}\right)^2+C'\left(\frac{1}{p}+\frac{1}{Np\Delta}+\Delta\right).$$

It remains to bound the risk on $\Omega_n^c$.

**Lemma 3.**
*If $n\Delta/\ln^2(n)\to\infty$ and $\mathscr{D}_n\leq cn\Delta/\ln^2(n)$, then*

$$\mathbb{P}(\Omega_n^c)\leq\frac{1}{n^2}.$$

The proof is done in Schmisser (2009b) section A.2.

Let us set $e_{k\Delta}=T_{k\Delta}-\sigma^2(\bar{Y}_{(k-1)\Delta})$, $e=(e_\Delta,\ldots,e_{n\Delta})^*$ and $\Pi_mT=\Pi_m\left(T_\Delta,\ldots,T_{n\Delta}\right)^*$, where $v*$ is the transpose of $v$ and $\Pi_m$ is the orthogonal projection of $\left(T_\Delta,\ldots,T_{n\Delta}\right)^*$ on $\left\{\left(t(\bar{Y}_0),\ldots,t(\bar{Y}_{(n-1)\Delta})\right)^*,t\in S_m\right\}$. We have

$$\begin{aligned}\left\|\hat{\sigma}_m^2-\sigma_A^2\right\|_n^2 &= \left\|\Pi_mT-\sigma_A^2\right\|_n^2\leq 2\left\|\sigma_A^2-\Pi_m\sigma_A^2\right\|_n^2+2\left\|\Pi_m\sigma_A^2-\Pi_mT\right\|_n^2\\ &\leq 2\left\|\sigma_A^2\right\|_n^2+2\left\|\Pi_me\right\|_n^2\leq 2\left\|\sigma_A^2\right\|_n^2+2\left\|e\right\|_n^2.\end{aligned}$$

Strict stationarity and the Cauchy-Schwartz inequality imply that

$$\mathbb{E}\left(\left\|\sigma_A^2\right\|_n^2\mathbb{1}_{\Omega_n^c}\right)\leq\sqrt{\mathbb{P}\left(\Omega_n^c\right)\mathbb{E}\left(\sigma_A^8(\bar{Y}_0)\right)}\leq\frac{c}{n}$$

and

$$\mathbb{E}\left(\left\|e\right\|_n^2\mathbb{1}_{\Omega_n^c}\right)\leq\sqrt{\mathbb{P}\left(\Omega_n^c\right)\mathbb{E}\left(e_\Delta^4\right)},$$

where $e_\Delta^4\leq c\left(A_\Delta^4+B_\Delta^4+R_\Delta^4+Z_\Delta^4+S_\Delta^4\right)$. Thanks to Lemma 1 and Lemma 2 **(iii)**, we have

$$\mathbb{E}\left(\left\|\hat{\sigma}_m^2-\sigma_A^2\right\|_n^2\mathbb{1}_{\Omega_n^c}\right)\leq\frac{c}{n}\left(\left(1+\frac{1}{p\Delta}\right)^2+\frac{1}{p}+\Delta\right).\tag{18}$$

Collecting terms, we obtain (as $np^2\Delta^2\to\infty$):

$$\mathbb{E}\left(\left\|\hat{\sigma}_m^2-\sigma_A^2\right\|_n^2\right)\leq 3\tilde{\pi}_1\left\|\sigma_m^2-\sigma_A^2\right\|_{L^2}^2+\frac{CD_m}{n}\left(\sigma_0^2+\frac{\tau^2}{p\Delta}\right)^2+C'\left(\Delta+\frac{1}{p}+\frac{1}{n}\left(1+\frac{1}{p\Delta}\right)^2\right).$$

## 5.3 Proof of Theorem 2

The bound of the risk on $\Omega_n^c$ is obtained exactly as for $m$ fixed, so we only bound the risk on $\Omega_n$. Let us consider the ball $\mathscr{B}_{m,m'}=\{t\in S_m+S_{m'},\|t\|_{\tilde{\pi}}\leq 1\}$. By definition of $\hat{m}$, for any $m\in\mathscr{M}_n$,

$$\left\|\hat{\sigma}_{\hat{m}}^2-\sigma_A^2\right\|_n^2+pen(\hat{m})\leq\left\|\hat{\sigma}_m^2-\sigma_A^2\right\|_n^2+pen(m).$$

Using Lemmas 1 and 2, we obtain

$$\begin{aligned}\mathbb{E}\left(\left\|\hat{\sigma}^2_{\hat{m}}-\sigma^2_A\right\|^2_n \mathbb{1}_{\Omega_n}\right) &\leq 3\tilde{\pi}_1\left\|\sigma^2_m-\sigma^2_A\right\|^2_{L^2}+2\mathbb{E}\left(pen(m)-pen(\hat{m})\right)\\ &+ 72\,\mathbb{E}\left(\sup_{t\in\mathscr{B}_{m,\hat{m}}}\left(\nu^2_n(t)+\rho^2_n(t)+\xi^2_n(t)\right)\mathbb{1}_{\Omega_n}\right)+C\left(\Delta+\frac{1}{p}+\frac{1}{n}\left(1+\frac{1}{p\Delta}\right)^2\right).\end{aligned}$$

It remains to bound the terms $\mathbb{E}\left(\sup_{t\in\mathscr{B}_{m,\hat{m}}}\nu^2_n(t)\right)$, $\mathbb{E}\left(\sup_{t\in\mathscr{B}_{m,\hat{m}}}\rho^2_n(t)\right)$ and $\mathbb{E}\left(\sup_{t\in\mathscr{B}_{m,\hat{m}}}\xi_n(t)\right)$. Let us introduce the functions

$$p_1(m,m')=\frac{\kappa_1\sigma_0^4\left(D_m+D_{m'}\right)}{n},\quad p_2(m,m')=\frac{\kappa_2\upsilon^4\left(D_m+D_{m'}\right)}{np^2\Delta^2},$$
$$p_3(m,m')=\frac{\kappa_3\upsilon^2\sigma_0^2\left(D_m+D_{m'}\right)}{np\Delta}$$

and the associated penalties

$$pen(m):=\frac{36\kappa_1\sigma_0^4D_m}{n}+\frac{36\kappa_2\upsilon^4D_m}{np^2\Delta^2}+\frac{36\kappa_3\upsilon^2\sigma_0^2D_m}{np\Delta}:=pen_1(m)+pen_2(m)+pen_3(m).$$

We have

$$\begin{aligned}E: &= \mathbb{E}\left(36\sup_{t\in\mathscr{B}_{m,\hat{m}}}\nu^2_n(t)+\left(pen_1(m)-pen_1(\hat{m})\right)\mathbb{1}_{\Omega_n}\right)\\ &\leq \mathbb{E}\left(36\sup_{t\in\mathscr{B}_{m,\hat{m}}}\left(\nu^2_n(t)-p_1(m,\hat{m})\mathbb{1}_{\Omega_n}\right)+\left(pen_1(m)-pen_1(\hat{m})+36p_1(m,\hat{m})\right)\mathbb{1}_{\Omega_n}\right).\end{aligned}$$

As

$$\left[\sup_{t\in\mathscr{B}_{m,\hat{m}}}\nu^2_n(t)-p_1(m,\hat{m})\right]_+\leq\sum_{m'\in\mathscr{M}_n}\left[\sup_{t\in\mathscr{B}_{m,m'}}\nu^2_n(t)-p_1(m,m')\right]_+,$$

we obtain

$$E\leq\mathbb{E}\left(36\sum_{m'\in\mathscr{M}_n}\left[\sup_{t\in\mathscr{B}_{m,m'}}\nu^2_n(t)-p_1(m,m')\right]_+\mathbb{1}_{\Omega_n}+2pen_1(m)\mathbb{1}_{\Omega_n}\right).\tag{19}$$

Likewise,

$$\begin{aligned}F: &= \mathbb{E}\left(36\sup_{t\in\mathscr{B}_{m,\hat{m}}}\rho^2_n(t)+\left(pen_2(m)-pen_2(\hat{m})\right)\mathbb{1}_{\Omega_n}\right)\\ &\leq \mathbb{E}\left(36\sum_{m'\in\mathscr{M}_n}\left[\sup_{t\in\mathscr{B}_{m,m'}}\rho^2_n(t)-p_2(m,m')\right]_+\mathbb{1}_{\Omega_n}+2pen_2(m)\mathbb{1}_{\Omega_n}\right)\end{aligned}\tag{20}$$

and

$$\begin{aligned}G: &= \mathbb{E}\left(36\sup_{t\in\mathscr{B}_{m,\hat{m}}}\xi^2_n(t)+\left(pen_3(m)-pen_3(\hat{m})\right)\mathbb{1}_{\Omega_n}\right)\\ &\leq \mathbb{E}\left(36\sum_{m'\in\mathscr{M}_n}\left[\sup_{t\in\mathscr{B}_{m,m'}}\xi^2_n(t)-p_3(m,m')\right]_+\mathbb{1}_{\Omega_n}+2pen_3(m)\mathbb{1}_{\Omega_n}\right)\end{aligned}\tag{21}$$

**Lemma 4.**
*There exists a universal constant $c_1$ such that, if Assumption 3 is satisfied:*

$$\mathbb{E}\left(\left|Z_{k\Delta}\right|^l \middle| \mathscr{G}_{k\Delta}\right) \leq \left(\sigma_0 c_1\right)^{2l} l^l$$

*where $\mathscr{G}_{k\Delta}$ is defined in* (15). *The variables $Z_{k\Delta}$ are $\mathscr{G}_{(k+2)\Delta}$-measurable.*

*Proof.* By definition,

$$Z_{k\Delta} = \frac{1}{\Delta}\left(\bar{J}^2_{k\Delta}(\sigma) - \frac{1}{p^2}\sum_{i,j}\int_{k\Delta+(i\vee j)\delta}^{(k+1)\Delta+(i\wedge j)\delta} \sigma^2\left(X_s\right) ds\right),$$

so

$$\begin{aligned}\mathbb{E}\left(\left|Z_{k\Delta}\right|^l\right) &\leq \frac{2^l}{\Delta^l}\left(\frac{1}{p}\sum_{j=1}^{p}\mathbb{E}\left\{\left(\int_{k\Delta+j\delta}^{(k+1)\Delta+j\delta}\sigma(X_s)dWs\right)^{2l} + \right.\right.\\ &+ \left.\left.\frac{1}{p^2}\sum_{i,j}\mathbb{E}\left[\left(\int_{k\Delta+(i\vee j)\delta}^{(k+1)\Delta+(i\wedge j)\delta}\sigma^2\left(X_s\right)\right)^l\right] ds\right\}\right)\end{aligned}$$

According to the Burkhölder-Davis-Gundy inequality with optimal constant, (Proposition 4.2 of Barlow and Yor (1982)), we have:

$$\mathbb{E}\left(\left|Z_{k\Delta}\right|^l\right) \leq \frac{2^{l+1}}{\Delta^l}\left(\frac{1}{p}\sum_{j=1}^{p}(2l)^l c^{2l}\sigma_0^{2l}\Delta^l\right) \leq 2\left(2\sigma_0 C\right)^{2l} l^l. \qquad \square$$

**Lemma 5.**
*If Assumptions 5 and 8 are satisfied, there exists a constant $c_2$ such that*

$$\mathbb{E}\left(\left|R_{k\Delta}\right|^l \middle| \mathscr{G}_{k\Delta}\right) \leq \frac{\left(\upsilon c_2\right)^{2l} l^l}{p^l\Delta^l}.$$

*The variables $R_{k\Delta}$ are $\mathscr{G}_{(k+2)\Delta}$-measurable.*

*Proof.* The variables $R_{k\Delta}$ are independent of $\mathscr{G}_{k\Delta}$, so $\mathbb{E}\left(\left|R_{k\Delta}\right|^l \middle| \mathscr{G}_{k\Delta}\right) = \mathbb{E}\left(\left|R_{k\Delta}\right|^l\right)$. We have

$$\begin{aligned}\mathbb{E}\left(\left|R_{k\Delta}\right|^l\right) &= \frac{1}{\Delta^l}\mathbb{E}\left[\left|\left(\bar{\varepsilon}_{k+1}-\bar{\varepsilon}_k\right)^2 - 2\tau^2/p\right|^l\right]\\ &\leq \frac{2^l}{\Delta^l}\left(2^{2l+1}\mathbb{E}\left(\bar{\varepsilon}_k^{2l}\right) + \frac{2^l\tau^{2l}}{p^l}\right)\\ &\leq \frac{4^l}{\Delta^l p^l}\left(2^{l+1}\mathbb{E}\left[\frac{1}{p^l}\left(\sum_{j=1}^{p}\varepsilon_{kp+j}\right)^{2l}\right] + \tau^{2l}\right).\end{aligned}$$

According to Assumption 8, $\tau^2 \leq \upsilon^2$. Moreover, as $\mathbb{E}\left(\varepsilon_{kp+j}^{2l+1}\right) = 0$ for any $l$, we can write:

$$\begin{aligned}\mathbb{E}\left[\left(\sum_{j=1}^{p}\varepsilon_{kp+j}\right)^{2l}\right] &= \sum_{1\leq j_1,\ldots,j_l\leq p}\mathbb{E}\left(\varepsilon_{kp+j_1}^2\ldots\varepsilon_{kp+j_l}^2\right)\\ &= \sum_{\sum a_j = l}\mathbb{E}\left(\varepsilon_{kp+1}^{2a_1}\ldots\varepsilon_{kp+p}^{2a_p}\right).\end{aligned}$$

According to Assumption 8,

$$\mathbb{E}\left[\left(\sum_{j=1}^{p}\varepsilon_{kp+j}\right)^{2l}\right] \leq \upsilon^{2l}\sum_{\sum a_j=l} a_1^{a_1}\dots a_p^{a_p} \leq \upsilon^{2l}\sum_{\sum a_j=l} l^l \leq p^l l^l \upsilon^{2l}.$$

So

$$\mathbb{E}\left(\left|R_{k\Delta}\right|^l\right) \leq c^l \frac{l^l \upsilon^{2l}}{\Delta^l p^l}$$

where $c$ is a constant.

□

**Lemma 6.**
*If Assumptions 3, 5 and 8 are satisfied:*

$$\mathbb{E}\left(S_{k\Delta}^{2l}\middle|\mathscr{G}_{k\Delta}\right) \leq \left(2c_1c_2\upsilon\sigma_0\right)^{2l}\frac{l^{2l}}{\Delta^l p^l}$$

*and*

$$\mathbb{E}\left(S_{k\Delta}^{2l+1}\middle|\mathscr{G}_{k\Delta}\right) = 0$$

*The random variables $S_{k\Delta}$ are $\mathscr{G}_{(k+2)\Delta}$-measurable.*

*Proof.* By definition,

$$S_{k\Delta} = \frac{2}{\Delta}\left(\bar{\varepsilon}_{k+1}-\bar{\varepsilon}_k\right)\bar{J}_k\left(\sigma\right).$$

So, $\mathbb{E}\left(S_{k\Delta}^{2l+1}\middle|\mathscr{G}_{k\Delta}\right) = 2^{2l+1}\Delta^{-l}\mathbb{E}\left[\left(\bar{\varepsilon}_{k+1}-\bar{\varepsilon}_k\right)^{2l+1}\right]\mathbb{E}\left[\left(\bar{J}_k\left(\sigma\right)\right)^{2l+1}\middle|\mathscr{G}_{k\Delta}\right] = 0$. Moreover, by independence:

$$\mathbb{E}\left(S_{k\Delta}^{2l}\middle|\mathscr{G}_{k\Delta}\right) = 2^{2l}\Delta^{-2l}\mathbb{E}\left[\left(\bar{\varepsilon}_{k+1}-\bar{\varepsilon}_k\right)^{2l}\right]\mathbb{E}\left[\left(\bar{J}_k\left(\sigma\right)\right)^{2l}\middle|\mathscr{G}_{k\Delta}\right]$$

and, thanks to Lemmas 4 and 5, we deduce:

$$\mathbb{E}\left(S_{k\Delta}^{2l}\right) \leq \left(2c_1c_2\right)^{2l}\frac{\left(\upsilon\sigma_0\right)^{2l}l^{2l}}{\Delta^l p^l}.$$

□

**Lemma 7.**
*Let us consider $f_n : S_{m,m'} := S_m + S_{m'} \to \mathbb{R}$ such that, for any function $t \in S_{m,m'}$,*

$$f_n(t) = \frac{1}{n}\sum_{k=1}^{n} U_{k\Delta}\, t\left(\bar{Y}_{(k-1)\Delta}\right)$$

*where $U_{k\Delta}$ are centred random variables and $\mathscr{G}_{(k+2)\Delta}$-measurable. If there exists a constant $\gamma$ such that*

$$\mathbb{E}\left(U_{k\Delta}^{l}\middle|\mathscr{G}_{k\Delta}\right) \leq \gamma^l l^l,$$

*then there exists a constant $\kappa > 0$ such that*

$$\mathbb{E}\left[\sup_{t\in\mathscr{B}_{m,m'}} f_n^2(t) - \frac{\kappa\gamma^2\left(D_m+D_{m'}\right)}{n}\right]_+ \leq \frac{\kappa\gamma^2 e^{-\left(D_m+D_{m'}\right)}}{n}$$

According to Lemmas 4-6, the processes $\nu_n(t)$, $\rho_n(t)$ and $\xi_n(t)$ satisfy the assumptions of Lemma 7 with respectively $\gamma = \sigma_0^2 c_1^2$, $\gamma = \upsilon^2 c_2^2/(p\Delta)$ and $\gamma = 2c_1c_2\upsilon\sigma_0/\sqrt{p\Delta}$. Then, there exist constants $\kappa_1$, $\kappa_2$ and $\kappa_3$ such that

$$\mathbb{E}\left[\sup_{t\in\mathscr{B}_{m,m'}} \nu_n^2(t) - \frac{\kappa_1\sigma_0^4\left(D_m + D_{m'}\right)}{n}\right]_+ \leq \frac{\kappa_1\sigma_0^4 e^{-(D_m+D_{m'})}}{n},$$

$$\mathbb{E}\left[\sup_{t\in\mathscr{B}_{m,m'}} \rho_n^2(t) - \frac{\kappa_2\upsilon^4\left(D_m + D_{m'}\right)}{np^2\Delta^2}\right]_+ \leq \frac{\kappa_2\upsilon^4 e^{-(D_m+D_{m'})}}{np^2\Delta^2}$$

and

$$\mathbb{E}\left[\sup_{t\in\mathscr{B}_{m,m'}} \xi_n^2(t) - \frac{\kappa_3\upsilon^2\sigma_0^2\left(D_m + D_{m'}\right)}{np\Delta}\right]_+ \leq \frac{\kappa_3\upsilon^2\sigma_0^2 e^{-(D_m+D_{m'})}}{np\Delta}.$$

As $\sum_{m'} e^{-D_{m'}} \leq 1$, according to (12), (13), (20) and (21), we have

$$E \leq 36\frac{\kappa_1\sigma_0^4}{n}, \qquad F \leq 36\frac{\kappa_2\upsilon^4}{np^2\Delta^2} \quad \text{and} \quad G \leq 36\frac{\kappa_3\upsilon^2\sigma_0^2}{np\Delta}.$$

So

$$\mathbb{E}\left(\left\|\hat{\sigma}_m^2 - \sigma_A^2\right\|_n^2\right) \leq C \inf_{m\in\mathscr{M}_n}\left(\left\|\hat{\sigma}_m^2 - \sigma_A^2\right\|_n^2 + pen(m)\right) + \frac{c}{n}\left(\tau^2 + \frac{\upsilon^2}{p\Delta}\right)^2 + C\Delta + \frac{C}{p}.$$

### 5.3.1 Proof of Lemma 7

First, we prove the lemma

**Lemma 8.**
*There exist constants $\alpha_1$ and $\alpha_2$ such that, if $\|t\|_\infty \leq c\zeta$,*

$$\mathbb{P}\left(f_n(t) \geq \eta, \|t\|_n^2 \leq \zeta^2\right) \leq 2\exp\left(-\frac{\eta^2 n}{(\alpha_1\gamma^2\zeta^2 + 2c\alpha_2\gamma\eta\zeta)}\right).$$

*Proof.* We can write

$$h := \mathbb{E}\left(\exp(\lambda t\left(\bar{Y}_{(k-1)\Delta}\right)U_{k\Delta}\middle|\mathscr{G}_{k\Delta}\right) = 1 + \sum_{l=2}^{+\infty}\frac{\lambda^l}{l!}t^l\left(\bar{Y}_{(k-1)\Delta}\right)\mathbb{E}\left(U_{k\Delta}^l\middle|\mathscr{G}_{k\Delta}\right),$$

so

$$h \leq 1 + \sum_{l=2}^{+\infty} c^l\gamma^l\lambda^l\frac{l^l}{l!}t^l\left(\bar{Y}_{(k-1)\Delta}\right).$$

As $\frac{l^l}{l!} \leq e^{l-1}$, we have

$$h \leq 1 + e^{-1}\sum_{l=2}^{+\infty}\left(c\gamma\lambda e\right)^l t^l\left(\bar{Y}_{(k-1)\Delta}\right).$$

So, for any $\lambda < \lambda_1 = 1/(c\gamma e\|t\|_\infty)$,

$$h \quad \leq \quad 1 + e^{-1}\frac{\left(c\gamma\lambda e\right)^2 t^2\left(\bar{Y}_{(k-1)\Delta}\right)}{1 - c\gamma\lambda e\|t\|_\infty},$$

and, setting $\alpha = (c\gamma)^2 e$, we obtain

$$h \leq 1 + \frac{\alpha\lambda^2 t^2\left(\bar{Y}_{(k-1)\Delta}\right)}{1-\lambda/\lambda_1} \leq \exp\left(\frac{\alpha\lambda^2 t^2\left(\bar{Y}_{(k-1)\Delta}\right)}{1-\lambda/\lambda_1}\right). \tag{22}$$

Let us set

$$f_l^{(1)}(t) = \sum_{k\leq l \text{ even}} U_{k\Delta} t\left(\bar{Y}_{(k-1)\Delta}\right),\ f_l^{(2)}(t) = \sum_{k\leq l \text{ odd}} U_{k\Delta} t\left(\bar{Y}_{(k-1)\Delta}\right).$$

According to the Tchebitchev inequality, for any $i \in \{1,2\}$ and $0 \leq \lambda < \lambda_1$:

$$\begin{aligned} F_1 &:= \mathbb{P}\left(f_n^{(i)}(t) \geq \frac{\eta n}{2}, \|t\|_n^2 \leq \zeta^2\right) \\ &\leq \mathbb{P}\left[\exp\left(\lambda f_n^{(i)}(t) - \frac{\alpha\lambda^2 n}{1-\lambda/\lambda_1}\|t\|_n^2\right) \geq \exp\left(\frac{\lambda\eta n}{3} - \frac{\alpha\lambda^2\zeta^2 n}{1-\lambda/\lambda_1}\right)\right] \\ &\leq \exp\left(-\frac{\lambda\eta n}{2} + \frac{\alpha\lambda^2\zeta^2 n}{1-\lambda/\lambda_1}\right)\mathbb{E}\left[\exp\left(\lambda f_n^{(i)}(t) - \frac{\alpha\lambda^2 n}{1-\lambda/\lambda_1}\|t\|_n^2\right)\right]. \end{aligned} \tag{23}$$

If $n$ is a multiple of 2, we have, as $U_{(n-2)\Delta}$ est $\mathscr{G}_{n\Delta}$-measurable,

$$\begin{aligned} F_2 &:= \mathbb{E}\left[\exp\left(\lambda f_n^{(1)}(t) - \frac{\alpha\lambda^2 n}{1-\lambda/\lambda_1}\|t\|_n^2\right)\right] \\ &= \mathbb{E}\left\{\exp\left(\lambda f_{n-2}^{(1)}(t) - \frac{\alpha\lambda^2}{1-\lambda/\lambda_1}\sum_{k=1}^{n-1} t^2\left(\bar{Y}_{(k-1)\Delta}\right)\right)\right. \\ &\quad \left.\mathbb{E}\left[\exp\left(\lambda U_{n\Delta} t\left(\bar{Y}_{(n-1)\Delta}\right) - \frac{\alpha\lambda^2}{1-\lambda/\lambda_1} t^2\left(\bar{Y}_{(n-1)\Delta}\right)\right)\middle| \mathscr{G}_{n\Delta}\right]\right\} \end{aligned}$$

So, according to (22):

$$\mathbb{E}\left[\exp\left(\lambda f_n^{(1)}(t) - \frac{\alpha\lambda^2 n}{1-\lambda/\lambda_1}\|t\|_n^2\right)\right] \leq \mathbb{E}\left[\exp\left(\lambda f_{n-2}^{(1)}(t) - \frac{\alpha\lambda^2}{1-\lambda/\lambda_1}\sum_{k=1}^{n-2} t^2\left(\bar{Y}_{(k-1)\Delta}\right)\right)\right].$$

By induction, we obtain

$$\mathbb{E}\left[\exp\left(\lambda f_n^{(1)}(t) - \frac{\alpha\lambda^2 n}{1-\lambda/\lambda_1}\|t\|_n^2\right)\right] \leq 1.$$

We can prove in the same way that $\mathbb{E}\left[\exp\left(\lambda f_n^{(2)}(t) - \frac{\alpha\lambda^2 n}{1-\lambda/\lambda_1}\|t\|_n^2\right)\right] \leq 1$. Moreover, as

$$\mathbb{P}\left(f_n(t) \geq \eta, \|t\|_n^2 \leq \zeta^2\right) \leq \sum_{i=1}^{2}\mathbb{P}\left(f_n^{(i)}(t) \geq \frac{\eta}{2}, \|t\|_n^2 \leq \zeta^2\right),$$

we obtain, thanks to (23),

$$\mathbb{P}\left(f_n(t) \geq cn, \|t\|_n^2 \leq c'\right) \leq 2\exp\left(-\frac{\lambda\eta n}{2} + \frac{\alpha\lambda^2\zeta^2 n}{1-\lambda/\lambda_1}\right).$$

Let us set $\lambda = \eta/\left(4\alpha\zeta^2 + \eta/\lambda_1\right)$. In that case, $\lambda < \lambda_1$ and

$$\mathbb{P}\left(f_n(t) \geq \eta, \|t\|_n^2 \leq \zeta^2\right) \leq 2\exp\left(-\frac{\eta^2 n}{4\left(4\alpha\zeta^2 + \eta/\lambda_1\right)}\right).$$

Substituting $\alpha$ and $\lambda_1$, we obtain the predicted result.

□

We complete the proof of Lemma 7. In this proof, $C$ is a constant which may vary from one line to another. The dimension of $S_{m,m'}$ is $D = \max(D_m, D_{m'})$. Let us consider an orthonormal (for the $L^2_{\tilde{\pi}}$ norm) basis $(\psi_\lambda)_{\lambda \in \Lambda_m}$ of $S_{m,m'}$ such that

$$\forall \lambda, \quad \text{card}\left(\{\lambda', \quad \|\psi_\lambda \psi_{\lambda'}\| \neq 0\}\right) \leq \phi_2 = 3(r-1).$$

(see Proposition 2). Let us set

$$\bar{r}_{m,m'} = \frac{1}{\sqrt{D}} \sup_{\beta \neq 0} \frac{\|\sum_\lambda \beta_\lambda \psi_\lambda\|_\infty}{|\beta|_\infty}.$$

It can easily be proved that

$$\left\| \sum_\lambda \beta_\lambda \psi_\lambda \right\|_\infty \leq 3(r-1)\,|\beta|_\infty \sup_\lambda \|\psi_\lambda\|_\infty \quad \text{and} \quad \|\psi_\lambda\|_\infty \leq \sqrt{D}\,\|\psi_\lambda\|_{L^2} \leq \tilde{\pi}_1 \sqrt{D}\,\|\psi_\lambda\|_{\tilde{\pi}}\,.$$

So

$$\bar{r}_{m,m'} \leq \bar{r} = 3(r-1)\tilde{\pi}_1.$$

The following result is exactly Lemma 9 in Barron *et al.* (1999).

*Result 1.* There exist $\delta_k$-lattices $T_k$ of $L^2_{\tilde{\pi}} \cap (S_m + S_{m'})$ such that

$$|T_k \cap \mathscr{B}_{m,m'}| \leq \left(5/\delta^k\right)^D$$

where $\delta_k = 2^{-k}/5$ . Let us denote by $p_k(u)$ the projection of $u$ over $T_k$. For any $u \in S_{m,m'}$, $\|u - p_k(u)\|_{\tilde{\pi}} \leq \delta_k$ and

$$\sup_{u \in p_k^{-1}(t)} \|u - t\|_\infty \leq \bar{r}_{m,m'}\delta_k \leq 3(r-1)\tilde{\pi}_1\delta_k.$$

Let us set $H_k = \ln(|T_k \cap \mathscr{B}_{,'}|)$. We have

$$H_k \leq D \ln(5/\delta_k) = D\left(k\ln(2) + \ln(5/\delta_0)\right) \leq C(k+1)D.$$

For any function $u \in \mathscr{B}_{m,m'}$, there exists a series $(u_k)_{k\geq 0}$ such that, for any $k$, $u_k \in T_k \cap \mathscr{B}_{m,m'}$, $\|u - u_k\|_{\tilde{\pi}} \leq \delta_k$ and $\|u - u_k\|_\infty \leq \bar{r}\delta_k$. Then, $u$ can be written as $u = u_0 + \sum_{k=1}^\infty (u_k - u_{k-1})$. Let us set

$$\mathscr{E}_k = \{u_k \in T_k \cap \mathscr{B}_{m,m'}, \quad \|u - u_k\|_{\tilde{\pi}} \leq \delta_k \quad \text{and} \quad \|u - u_k\|_\infty \leq c\delta_k\}\,.$$

Of course, $|\mathscr{E}_k| \leq H_k$. Let us set $\mathbb{P}_n(.) = \mathbb{P}\left(. \cap \Omega_n\right)$ and consider a series $(\eta_k)_{k\geq 0}$ such that $\eta_0 + \sum_{k=1}^\infty \eta_k \leq \eta$. We have

$$\begin{aligned}
\mathbb{P}_n\left(\sup_{u \in \mathscr{B}_{m,m'}} f_n(u) > \eta\right) &\leq \mathbb{P}_n\left(\exists (u_k),\ f_n(u_0) + \sum_{k=1}^\infty f_n(u_k - u_{k-1}) > \eta_0 + \sum_{k=1}^\infty \eta_k\right) \\
&\leq \sum_{u_0 \in \mathscr{E}_0} \mathbb{P}_n\left(f_n(u_0) > \eta_0\right) + \sum_{k=1}^\infty \sum_{u_k \in \mathscr{E}_k} \mathbb{P}_n\left(f_n(u_k - u_{k-1}) > \eta_0\right) \\
&\leq \mathbb{P}_1 + \sum_{k=1}^\infty \mathbb{P}_{2,k}.
\end{aligned}$$

As $u_0 \in T_0$, $\|u_0\|_{\tilde{\pi}} \leq 1$ and $\|u_0\|_\infty \leq \bar{r}$. The norms $\|.\|_n$ and $\|.\|_{\tilde{\pi}}$ are equivalent and $\|u_0\|_n^2 \leq 3/2\|u_0\|_{\tilde{\pi}}^2 \leq 3\delta_0/2$. Then

$$\mathbb{P}_n\left(f_n(u_0) > \eta_0\right) = \mathbb{P}_n\left(f_n(u_0) > \eta_0,\ \|u_0\|_n^2 \leq 3\delta_0/2\right).$$

According to Lemma 8, there exist constants $c_1'$ and $c_2'$ such that

$$\mathbb{P}_n\left(f_n(u_0) > \eta_0\right) \leq \exp\left(-\frac{n\eta_0^2}{c_1'\gamma^2 + 2c_2'\gamma\eta_0}\right).$$

Let us fix $x_0$ such that $\eta_0 = \gamma\left(\sqrt{c_1' x_0} + c_2' x_0\right)$. We have

$$x_0 \leq \frac{\eta_0^2}{c_1'\gamma^2 + 2c_2'\gamma\eta_0}$$

and

$$\mathbb{P}_n\left(f_n(u_0) > \eta_0\right) \leq \exp\left(-nx_0\right).$$

Hence,

$$\mathbb{P}_1 \leq \exp\left(H_0 - nx_0\right).$$

Likewise, $\|u_k - u_{k-1}\|_{\tilde{\pi}} \leq \|u - u_{k-1}\|_{\tilde{\pi}} + \|u - u_k\|_{\tilde{\pi}} \leq (2^{-(k-1)} + 2^{-k})\delta_0 = 2^{-k}3\delta_0$, so $\|u_k - u_{k-1}\|_n^2 \leq 14\delta_0^2\, 2^{-2k}$. Moreover, $\|u_k - u_{k-1}\|_\infty \leq (\delta_{k-1} + \delta_k)\,\bar{r} = 3.2^{-k}\delta_0\bar{r}$. According to Lemma 7, there exist constants $c_3$ and $c_4$ such that:

$$\begin{aligned}\mathbb{P}_n\left(f_n(u_k - u_{k-1}) > \eta_k\right) &= \mathbb{P}_n\left(f_n(u_k - u_{k-1}) > \eta_k,\ \|u_k - u_{k-1}\|_n^2 \leq C2^{-2k}\right)\\ &\leq \exp\left(-\frac{n\eta_k^2}{c_3\gamma^2 2^{-2k} + 2c_4\gamma 2^{-k}}\right).\end{aligned}$$

Fix $x_k$ such that $\eta_k = 2^{-k}\left(\sqrt{c_3 x_k} + c_4 x_k\right)$. We have

$$x_k \leq \frac{\eta_k^2}{c_3\gamma^2 2^{-2k} + 2c_4\gamma 2^{-k}}$$

and

$$\mathbb{P}_n\left(f_n(u_k - u_{k-1}) > \eta_k\right) \leq \exp\left(-nx_k\right).$$

Hence, $\mathbb{P}_{2,k} \leq \exp\left(H_{k-1} + H_k - nx_k\right)$ and $\mathbb{P}_2 = \sum_{k=1}^\infty \mathbb{P}_{2,k} \leq \sum_{k=1}^\infty \exp\left(H_{k-1} + H_k - nx_k\right)$. Let us fix $\tau > 0$ and the series $(x_k)$ (so the series $(\eta_k)$) such that

$$\begin{cases} nx_0 = H_0 + D + \tau \\ nx_k = H_{k-1} + H_k + (k+1)D + \tau \end{cases}$$

We have

$$\mathbb{P}_n\left(\sup_{u \in \mathscr{B}_{m,m'}} f_n(u) > \eta\right) \leq e^{-\tau}\sum_{k=1}^\infty \exp\left(-(k+1)D\right) \leq e^{-\tau}\frac{e^{-D}}{1 - e^{-D}} \leq 1.6e^{-D}e^{-\tau}.$$

It remains to compute $\eta^2$. Here, $C$ denotes a constant which can vary from one line to another but which is independent of $\gamma$. We have

$$\eta = \sum_{k=0}^\infty \eta_k \quad \leq \quad C\gamma\left(\sum_{k=0}^\infty 2^{-k+1}\left(\sqrt{x}_k + x_k\right)\right).$$

We recall that $H_k = C(k+1)D$. Then, $nx_k = C(3k+2)D + \tau$ and

$$\sum_{k=0}^{\infty} 2^{-k+1} x_k \leq \frac{1}{n} \sum_{k=0}^{\infty} 2^{-k} (C(3k+2)D + \tau) \leq C \frac{D+\tau}{n}.$$

Moreover,

$$\sum_{k=0}^{\infty} 2^{-k+1} \sqrt{x_k} \leq C \frac{\sqrt{D} + \sqrt{\tau}}{\sqrt{n}}.$$

As $D/n \leq 1$:

$$\eta^2 \leq C\gamma^2 \left( \frac{D}{n} + 2\frac{\tau}{n} + \frac{\tau^2}{n^2} \right).$$

So

$$\mathbb{P}_n \left( \sup_{u \in \mathscr{B}_{m,m'}} f_n^2(u) > C\gamma^2 \left( \frac{D}{n} + 2\frac{\tau}{n} + \frac{\tau^2}{n^2} \right) \right) \leq 1.6 e^{-D-\tau}. \tag{24}$$

Besides,

$$\begin{aligned} E &:= \mathbb{E} \left( \left[ \sup_{u \in \mathscr{B}_{m,m'}} f_n^2(u) - C\gamma^2 \frac{D}{n} \right]_+ \mathbb{1}_{\Omega_n} \right) \\ &= \int_0^{\infty} \mathbb{P}_n \left( \sup_{u \in \mathscr{B}_{m,m'}} f_n^2(u) > C\gamma^2 \frac{D}{n} + \tau \right) d\tau \end{aligned}$$

Substituting $\tau = C\gamma^2 \left( 2y/n + y^2/n^2 \right)$, we find that

$$E = C\gamma^2 \int_0^n \mathbb{P}_n \left( \sup_{u \in \mathscr{B}_{m,m'}} f_n^2(u) > C\gamma^2 \left( \frac{D}{n} + 2\frac{y}{n} + \frac{y^2}{n^2} \right) \right) \left( \frac{2}{n} + \frac{2y}{n^2} \right) dy.$$

According to (24),

$$\begin{aligned} E &= 2C\gamma^2 e^{-D} \left( \frac{1}{n} \int_0^{\infty} e^{-y} dy + \frac{1}{n^2} \int_0^{\infty} y e^{-y} dy \right) \\ &\leq \frac{C\gamma^2}{n} e^{-D}. \end{aligned}$$

which ends the proof of Lemma 7.

## 5.4 Proof of theorem 3

It is now assumed that $p\Delta \to \infty$. Let us set $e_{n,m} = E_n \left( \hat{\sigma}_m^2 - \sigma_m^2 \right) + \rho_n \left( \hat{\sigma}_m^2 - \sigma_m^2 \right)$. One can easily obtain that

$$\left\| \hat{\sigma}_m^2 - \sigma_A^2 \right\|_n^2 \leq \left\| \sigma_m^2 - \sigma_A^2 \right\|_n^2 + 2\nu_n \left( \hat{\sigma}_m^2 - \sigma_m^2 \right) + 2e_{n,m}.$$

We obtain, on $\Omega_n$,

$$\left\| \hat{\sigma}_m^2 - \sigma_A^2 \right\|_n^2 \leq 3 \left\| \hat{\sigma}_m^2 - \sigma_m^2 \right\|_n^2 + 24 \sup_{t \in \mathscr{B}_m} \nu_n^2(t) + \frac{96}{n} \sum_{k=1}^{n} \left( A_{k\Delta}^2 + B_{k\Delta}^2 + R_{k\Delta}^2 + S_{k\Delta}^2 \right).$$

According to Lemma 1 and inequality (17),

$$\mathbb{E}\left(\left\|\hat{\sigma}_m^2-\sigma_A^2\right\|_n^2\mathbb{1}_{\Omega_n}\right)\leq\frac{c\sigma_0^4D_m}{n}+C\left(\frac{1}{p^2\Delta^2}+\frac{1}{p\Delta}+\frac{1}{p}+\Delta\right).$$

Thanks to (18), and, as $p\Delta\to\infty$, we have, on $\Omega_n^c$,

$$\mathbb{E}\left(\left\|\hat{\sigma}_m^2-\sigma_A^2\right\|_n^2\mathbb{1}_{\Omega_n^c}\right)\leq\frac{C}{n}.$$

Collecting terms, we find that

$$\mathbb{E}\left(\left\|\hat{\sigma}_m^2-\sigma_A^2\right\|_n^2\mathbb{1}_{\Omega_n}\right)\leq\frac{c\sigma_0^4D_m}{n}+C\left(\frac{1}{p^2\Delta^2}+\frac{1}{p\Delta}+\frac{1}{p}+\Delta+\frac{1}{n}\right).$$

Let us set

$$p(m,m')=p_1(m,m')=\frac{\kappa\sigma_0^4}{n}\left(D_m+D_{m'}\right),\qquad pen(m)=\frac{24\kappa\sigma_0^4D_m}{n}$$

and

$$E'=\mathbb{E}\left[\left(24\sup_{t\in\mathscr{B}_{m,\hat{m}}}\nu_n^2(t)+(pen(m)-pen(\hat{m}))\right)\mathbb{1}_{\Omega_n}\right].$$

According to the previous proof, there exists $\kappa$ such that:

$$E'\leq 24\kappa\frac{\sigma_0^4}{n}+2pen(m),$$

then

$$\mathbb{E}\left(\left\|\hat{\sigma}_m^2-\sigma_A^2\right\|_n^2\right)\leq C\inf_{m\in\mathscr{M}_n}\left(\left\|\sigma_m^2-\sigma_A^2\right\|_{L^2}^2+pen(m)\right)+C\left(\frac{1}{n}+\frac{1}{p^2\Delta^2}+\frac{1}{p\Delta}+\frac{1}{p}+\Delta\right).$$

This inequality is minimum when $\frac{1}{p\Delta}=\Delta$, (or when $p=\Delta^{-2}=\delta^{-2/3}$), and in this case:

$$\mathbb{E}\left(\left\|\hat{\sigma}_m^2-\sigma_A^2\right\|_n^2\right)\leq C\inf_{m\in\mathscr{M}_n}\left(\left\|\sigma_m^2-\sigma_A^2\right\|_{L^2}^2+pen(m)\right)+C\left(\frac{1}{n}+\delta^{1/3}\right).$$

# A Proofs of Lemmas 2 and 1

We recall here the Rosenthal inequality (see Hall and Heyde (1980) theorem 2.12 p.23).

**The Rosenthal inequality**

Let $(\eta_1,\ldots,\eta_n)$ be centred and independent variables such that $\mathbb{E}\left(|\eta_i|^p\right)<\infty$. Then, there exists a positive constant $r_p$ such that

$$\mathbb{E}\left(\left|\sum_{i=1}^n\eta_i\right|^p\right)\leq r_p\left[\sum_{i=1}^n\mathbb{E}\left|\eta_i\right|^p+\left(\sum_{i=1}^n\mathbb{E}\left(\eta_i^2\right)\right)^{p/2}\right].$$

## A.1 Proof of Lemma 1

**(i)** According to the Burkhölder inequality and equations (6) and (9):

$$\begin{aligned}\mathbb{E}\left(Z_{k\Delta}^2\middle|\mathscr{G}_{k\Delta}\right) &\leq \frac{8}{\Delta^2}\mathbb{E}\left\{\left[\frac{1}{p}\sum_{j=1}^{p}\left(\int_{k\Delta+j\delta}^{(k+1)\Delta+j\delta}\sigma(X_s)dW_s\right)^4+\right.\right.\\ &+ \left.\left.\frac{1}{p^2}\sum_{i,j}\left(\int_{k\Delta+(i\vee j)\delta}^{(k+1)\Delta+(i\wedge j)\delta}\sigma^2(X_s)ds\right)^2\right]\middle|\mathscr{G}_{k\Delta}\right\}\\ &\leq c\sigma_0^4.\end{aligned}$$

and $\mathbb{E}\left(Z_{k\Delta}^4\right)\leq c\sigma_0^8$.

**(ii)** According to the Rosenthal inequality and equation (7),

$$\mathbb{E}\left(R_{k\Delta}^2\right)\leq\frac{1}{\Delta^2}\mathbb{E}\left(\left(\bar{\varepsilon}_{k+1}-\bar{\varepsilon}_k\right)^4\right)\leq\frac{c\tau^4}{p^2\Delta^2}+\frac{c\tau_4^4}{p^3\Delta^2}$$

and $\mathbb{E}\left(R_{k\Delta}^4\right)\leq c\tau^8/p^4\Delta^4+c\tau_8^8/(p^7\Delta^4)$.

**(iii)** As the random variables $(\bar{\varepsilon}_k)$ are independent of $(X_t)_{t\geq 0}$,

$$\mathbb{E}\left[\bar{J}_k^2(\sigma)\middle|\mathscr{G}_{k\Delta}\right]=\mathbb{E}\left[\left(\int_{k\Delta+\delta}^{(k+2)\Delta}H_k^\sigma(s)dW_s\right)^2\middle|\mathscr{G}_{k\Delta}\right]\leq c\Delta\sigma_0^2.$$

and

$$\mathbb{E}\left[\left(S_{k\Delta}\right)^2\right]=\frac{1}{\Delta^2}\mathbb{E}\left[\left(\bar{\varepsilon}_k-\bar{\varepsilon}_{k-1}\right)^2\right]\mathbb{E}\left[\bar{J}_k^2(\sigma)\right]$$

where $S_{k\Delta}$ is defined by (8) and $\bar{J}_k(t)$ by (9). As $E\left[\left(\bar{\varepsilon}_k-\bar{\varepsilon}_{k-1}\right)^2\right]=\tau^2/p$,

$$\mathbb{E}\left[\left(S_{k\Delta}\right)^2\right]\leq c\tau^2\sigma_0^2/(p\Delta).$$

We can prove in the same way that $\mathbb{E}\left[\left(S_{k\Delta}\right)^4\right]\leq c\tau^4\sigma_0^4/(p^2\Delta^2)$.

**(iv)** By the Cauchy-Schwartz inequality, stationarity and (10)

$$\begin{aligned}\mathbb{E}\left(\left(A_{k\Delta}^{(1)}\right)^2\right) &\leq \frac{1}{p\Delta^2}\sum_{j=1}^{p}\left(\int_{k\Delta+j\delta}^{(k+1)\Delta+j\delta}b(X_s)ds\right)^4\\ &\leq \frac{\Delta}{p}\sum_{j=1}^{p}\int_{k\Delta+j\delta}^{(k+1)\Delta+j\delta}b^4(X_s)ds\\ &\leq \Delta^2\mathbb{E}\left(b^4\left(X_0\right)\right).\end{aligned}$$

Likewise, $\mathbb{E}\left(\left(A_{k\Delta}^{(1)}\right)^4\right)\leq\Delta^4\mathbb{E}\left(b^8(X_0)\right)$. Moreover,

$$\mathbb{E}\left(\left(A_{k\Delta}^{(2)}\right)^2\right)\leq\frac{4}{\Delta^2}\sqrt{\mathbb{E}\left(\bar{I}_{k\Delta}^4\right)}\sqrt{\mathbb{E}\left(\bar{J}_{k\Delta}^4\right)}\leq 4\sigma_0^2\sqrt{\mathbb{E}\left(b^4(X_0)\right)}\Delta$$

and $\mathbb{E}\left(\left(A_{k\Delta}^{(2)}\right)^4\right) \leq 16\sigma_0^4\sqrt{\mathbb{E}\left(b^8(X_0)\right)}\Delta^2$. According to Proposition 1, $\mathbb{E}\left[\left(\sigma^2\left(X_s\right)-\sigma^2\left(\bar{X}_{k\Delta}\right)\right)^2\right] \leq c\Delta$ if $s \in [k\Delta,(k+2)\Delta]$. So

$$\begin{aligned}
\mathbb{E}\left(\left(A_{k\Delta}^{(3)}\right)^2\right) &\leq \frac{1}{\Delta^2}\mathbb{E}\left[\left(\frac{1}{p^2}\sum_{i,j}\int_{k\Delta+(i\vee j)\delta}^{(k+1)\Delta+(i\wedge j)\delta}\sigma^2(X_s)-\sigma^2(\bar{X}_{(k-1)\Delta})ds\right)^2\right] \\
&\leq \frac{1}{p^2\Delta^2}\sum_{i,j}\mathbb{E}\left(\int_{k\Delta+j\delta}^{(k+1)\Delta+j\delta}\sigma^2(X_s)-\sigma^2(\bar{X}_{(k-1)\Delta})ds\right)^2 \\
&\leq \frac{1}{p^2\Delta}\sum_{i,j}\int_{k\Delta+(i\vee j)\delta}^{(k+1)\Delta+(i\wedge j)\delta}\mathbb{E}\left(\sigma^2(X_s)-\sigma^2(\bar{X}_{(k-1)\Delta})\right)^2 ds \\
&\leq c\Delta.
\end{aligned}$$

Likewise, $\mathbb{E}\left(\left(A_{k\Delta}^{(3)}\right)^4\right) \leq c\Delta^2$.

**(v)** According to (11), and as random variables $(\bar{\varepsilon}_k)$ are independent of $(X_t)_{t\geq 0}$,

$$\mathbb{E}\left(\left(B_{k\Delta}^{(1)}\right)^2\right) = \frac{4}{\Delta^2}\mathbb{E}\left(\left(\bar{\varepsilon}_{k+1}-\bar{\varepsilon}_k\right)^2\right)\mathbb{E}\left(\bar{I}_{k\Delta}^2\right) \leq c\tau^2/p.$$

In the same way, it can be proved that $\mathbb{E}\left(\left(B_{k\Delta}^{(1)}\right)^4\right) \leq c\tau^4/p^2$. Moreover, and $\sigma$ is Lipschitz,

$$\mathbb{E}\left(\left(B_{k\Delta}^{(2)}\right)^2\right) \leq \mathbb{E}\left(\left(\sigma^2(\bar{X}_{k\Delta})-\sigma^2(\bar{X}_\Delta+\bar{\varepsilon})\right)^2\right) \leq c\sigma_L\sigma_0\mathbb{E}\left(\bar{\varepsilon}^2\right) \leq \frac{c\sigma_L\sigma_0\tau^4}{p}$$

According to the Rosenthal inequality, $\mathbb{E}\left(\left(B_{k\Delta}^{(2)}\right)^4\right) \leq c\sigma_L\sigma_0^3\mathbb{E}\left(\bar{\varepsilon}^4\right) \leq c\sigma_L\sigma_0^3\tau^8/p^2$. Moreover,

$$\begin{aligned}
\sum_{i,j}\int_{k\Delta+(i\vee j)\delta}^{(k+1)\Delta+(i\wedge j)\delta} ds &= \sum_{i,j} i\wedge j+(p-i\vee j) = 2\sum_{i<j} i+p-j+\sum_i p \\
&= \frac{2}{3}p^2+O\left(p\right).
\end{aligned}$$

So according to (11), we have

$$\mathbb{E}\left(\left(B_{k\Delta}^{(3)}\right)^2\right) \leq \frac{c\sigma_0^4}{p^2} \qquad \text{and} \qquad \mathbb{E}\left(\left(B_{k\Delta}^{(3)}\right)^4\right) \leq \frac{c\sigma_0^8}{p^4}.$$

## A.2 Proof of Lemma 2

**(i)** According to (4) and the Itô formula,

$$\mathbb{E}\left(\hat{\tau}^2-\tau^2\right) = \frac{1}{2n'}\sum_{k=1}^{n}\mathbb{E}\left[\left(\left(X_{(k+1)\delta}-X_{k\delta}\right)+\left(\varepsilon_{k+1}-\varepsilon_k\right)\right)^2\right]-\tau^2. \tag{25}$$

The process $(X_t)$ is stationary, the random variables $(\varepsilon_k)$ have same law and are independent of $(X_t)$, so

$$2\mathbb{E}\left(\hat{\tau}^2-\tau^2\right)=\mathbb{E}\left[\left(X_\delta-X_0\right)^2\right]+\mathbb{E}\left[\left(\varepsilon_{k+1}-\varepsilon_k\right)^2\right]-2\tau^2.$$

Moreover, $\mathbb{E}\left[\left(\varepsilon_{k+1}-\varepsilon_k\right)^2\right]=2\tau^2$ and, by stationarity:

$$\mathbb{E}\left[\left(X_\delta-X_0\right)^2\right]\leq\mathbb{E}\left[\left(\int_0^\delta b(X_s)ds\right)^2+\left(\int_0^\delta\sigma(X_s)dW_s\right)^2\right]\leq\delta^2\mathbb{E}\left(b(X_0)\right)+\delta\mathbb{E}\left(\sigma^2(X_0)\right).$$

**(ii)** According to (25),

$$\begin{aligned}\mathbb{E}\left[\left(\hat{\tau}^2-\tau^2\right)^2\right] &= \mathbb{E}\left[\left(\frac{1}{4n'}\sum_{k=1}^{n'}\left(\varepsilon_{k+1}-\varepsilon_k+X_{(k+1)\delta}-X_{k\delta}\right)^2-2\tau^2\right)^2\right]\\ &= C_1+C_2+C_3+C_4+C_5+C_6.\end{aligned}$$

where

$$\begin{aligned}C_1 &= \frac{1}{4N^2}\left(\sum_{k=1}^{N}\left(\varepsilon_{k+1}-\varepsilon_k\right)^2-2\tau^2\right)^2\\ C_2 &= \frac{1}{N^2}\left(\sum_{k=1}^{N}\left(\varepsilon_{k+1}-\varepsilon_k\right)^2-2\tau^2\right)\left(\sum_{k=1}^{N}\left(\varepsilon_{k+1}-\varepsilon_k\right)\left(X_{(k+1)\delta}-X_{k\delta}\right)\right)\\ C_3 &= \frac{1}{N^2}\left(\sum_{k=1}^{N}\left(X_{(k+1)\delta}-X_{k\delta}\right)^2\right)\left(\sum_{k=1}^{N}\left(\varepsilon_{k+1}-\varepsilon_k\right)\left(X_{(k+1)\delta}-X_{k\delta}\right)\right)\\ C_4 &= \frac{1}{N^2}\left(\sum_{k=1}^{N}\left(\varepsilon_{k+1}-\varepsilon_k\right)\left(X_{(k+1)\delta}-X_{k\delta}\right)\right)^2\\ C_5 &= \frac{1}{4N^2}\left(\sum_{k=1}^{n}\left(X_{(k+1)\delta}-X_{k\delta}\right)^2\right)^2\\ C_6 &= \frac{1}{2N^2}\left(\sum_{k=1}^{n}\left(X_{(k+1)\delta}-X_{k\delta}\right)^2\right)\left(\sum_{k=1}^{n}\left(\varepsilon_{k+1}-\varepsilon_k\right)^2-\tau^2\right).\end{aligned}$$

We recall that the random variables $(\varepsilon_k)$ are independent of $(X_t)$ and, for $|k-l|\geq 2$, $\varepsilon_{k+1}-\varepsilon_k$ is independent of $\varepsilon_{l+1}-\varepsilon_l$. As $\mathbb{E}\left[\left(\varepsilon_{k+1}-\varepsilon_k\right)^2-2\tau^2\right]=0$, we find that

$$\begin{aligned}\mathbb{E}\left(C_1\right) &= \frac{1}{4N^2}\sum_{|k-l|\leq 1}\left[\left(\varepsilon_{k+1}-\varepsilon_k\right)^2-2\tau^2\right]\left[\left(\varepsilon_{l+1}-\varepsilon_l\right)^2-2\tau^2\right]\\ &\leq \frac{3c}{4N}\left(\tau_4+\tau^4\right).\end{aligned}$$

As the random variables $\bar{\varepsilon}_k$ are independent of $(X_t)_{t\geq 0}$, $\mathbb{E}\left(C_2\right)=\mathbb{E}\left(C_3\right)=0$. Moreover,

$$\mathbb{E}\left(C_4\right)=\frac{1}{N^2}\sum_{|k-l|\leq 1}\left(\varepsilon_{k+1}-\varepsilon_k\right)\left(X_{(k+1)\delta}-X_{k\delta}\right)\left(\varepsilon_{l+1}-\varepsilon_l\right)\left(X_{(l+1)\delta}-X_{l\delta}\right)$$

and

$$\begin{aligned}|\mathbb{E}(C_4)| &\leq \frac{1}{N^2}\sum_{|k-l|\leq 1}\mathbb{E}\left[\left(\varepsilon_{k+1}-\varepsilon_k\right)^2\right]\mathbb{E}\left[\left(X_{(k+1)\delta}-X_{k\delta}\right)^2\right] \\ &\leq \frac{3\delta\sigma_0^2\tau^2}{N}.\end{aligned}$$

According to proposition 1,

$$\mathbb{E}(C_5) \leq 2c\delta^2.$$

Moreover, by independence, $\mathbb{E}(C_6) = 0$.

**(iii)** We can prove in the same way that

$$\mathbb{E}\left[\left(\hat{\tau}^2-\tau^2\right)^4\right] \leq c\left(\frac{\tau^4}{N^2}+\delta^4\right).$$

# References


Barlow, M.T. and Yor, M. (1982) Semimartingale inequalities via the Garsia-Rodemich-Rumsey lemma, and applications to local times. *J. Funct. Anal.*, 49 (2) pp. 198–229.

Barron, A., Birgé, L. and Massart, P. (1999) Risk bounds for model selection via penalization. *Probab. Theory Related Fields*, 113 (3) pp. 301–413.

Comte, F. and Rozenholc, Y. (2002) Adaptive estimation of mean and volatility functions in (auto-)regressive models. *Stochastic Process. Appl.*, 97 (1) pp. 111–145.

Comte, F. and Rozenholc, Y. (2004) A new algorithm for fixed design regression and denoising. *Ann. Inst. Statist. Math.*, 56 (3) pp. 449–473.

Comte, F., Genon-Catalot, V. and Rozenholc, Y. (2007) Penalized nonparametric mean square estimation of the coefficients of diffusion processes. *Bernoulli*, 13 (2) pp. 514–543.

Gloter, A. (2000) Discrete sampling of an integrated diffusion process and parameter estimation of the diffusion coefficient. *ESAIM Probab. Statist.*, 4 pp. 205–227 (electronic).

Gloter, A. and Jacod, J. (2001) Diffusions with measurement errors. II. Optimal estimators. *ESAIM Probab. Statist.*, 5 pp. 243–260 (electronic).

Hall, P. and Heyde, C.C. (1980) *Martingale limit theory and its application*. Academic Press Inc. [Harcourt Brace Jovanovich Publishers], New York. Probability and Mathematical Statistics.

Jacod, J., Li, Y., Mykland, P.A., Podolskij, M. and Vetter, M. (2009) Microstructure noise in the continuous case: The pre-averaging approach. *Stochastic Processes and their Applications*.

Meyer, Y. (1990) *Ondelettes et opérateurs. I*. Actualités Mathématiques. [Current Mathematical Topics]. Hermann, Paris. Ondelettes. [Wavelets].

Pardoux, E. and Veretennikov, A.Y. (2001) On the Poisson equation and diffusion approximation. I. *Ann. Probab.*, 29 (3) pp. 1061–1085.

Rosenbaum, M. (2007) *étude de quelques problèmes d'estimation statistique en finance*. Ph.D. thesis, Université Paris-Est et CREST.

Schmisser, E. (2009a) Penalized nonparametric drift estimation for a multidimensional diffusion process. Preprint 2009-02, MAP5, Université Paris Descartes.

Schmisser, E. (2009b) Nonparametric drift estimation from noisy data. Preprint 2009-21, MAP5, Université Paris Descartes.

Zhang, L., Mykland, P.A. and Aït-Sahalia, Y. (2005) A tale of two time scales: determining integrated volatility with noisy high-frequency data. *J. Amer. Statist. Assoc.*, 100 (472) pp. 1394–1411.

Table 1: Model 1: Ornstein-Uhlenbeck

$$\sigma(x) = 1.$$

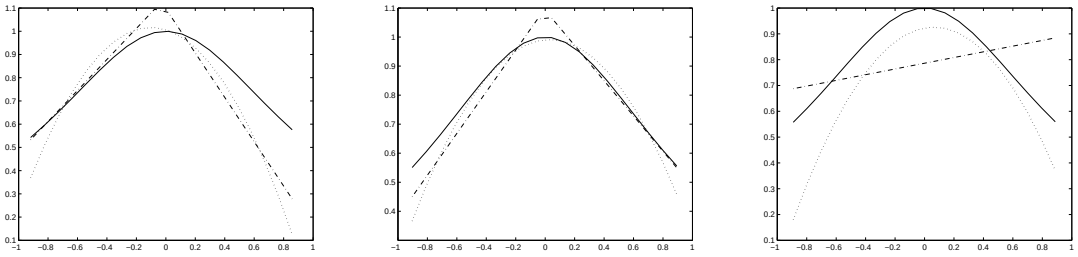
Estimation over $[-1, 1]$.
*ris*: empirical mean, *oracle*: oracle

| | | Gaussian noise | | Uniform noise | | Laplace noise | |
|---|---|---|---|---|---|---|---|
| $\Delta$ | $n$ | *risk* | *oracle* | *risk* | *oracle* | *risk* | *oracle* |
| $3.10^{-3}$ | $10^4$ | **0.022** | 2.78 | **0.0029** | 1.72 | 0.062 | 1.98 |
| $10^{-2}$ | $10^4$ | **0.022** | 2.79 | **0.0030** | 1.26 | **0.043** | 1.44 |
| $3.10^{-2}$ | $10^4$ | 0.027 | 2.33 | 0.0060 | 1.29 | 0.075 | 1.68 |
| $10^{-1}$ | $10^4$ | 0.032 | **1.23** | 0.023 | 1.20 | 0.064 | 1.33 |
| $3.10^{-3}$ | $10^3$ | 0.16 | 1.84 | 0.032 | 1.73 | 0.58 | 1.80 |
| $10^{-2}$ | $10^3$ | 0.22 | 3.52 | 0.034 | 1.46 | 0.51 | 2.08 |
| $3.10^{-2}$ | $10^3$ | 0.32 | 3.35 | 0.026 | 1.71 | 1.03 | 2.59 |
| $10^{-1}$ | $10^3$ | 0.11 | 1.56 | 0.015 | **1.09** | 0.52 | 1.86 |
| $10^{-2}$ | $10^2$ | 0.85 | 1.25 | 0.22 | 1.49 | 3.01 | **1.20** |
| $3.10^{-2}$ | $10^2$ | 1.43 | 1.32 | 0.35 | 1.70 | 4.31 | 1.82 |
| $10^{-1}$ | $10^2$ | 1.03 | 1.94 | 0.17 | 1.50 | 4.44 | 1.67 |

Figure 1: Model 3

$$\sigma^2(x) = \frac{1}{1+x^2}$$

Gaussian noise | Uniform noise | Laplace noise

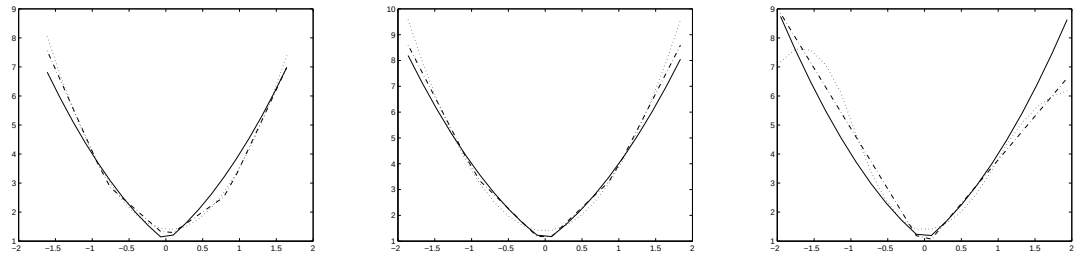


_ : true drift
. : estimated drift for $r = 1$
. : estimated drift for $r = 2$

Figure 2: Model 4

$$\sigma^2(x) = (1+|x|)^2$$

Gaussian noise | Uniform noise | Laplace noise



_ : true drift
. : estimated drift for $r = 1$
. : estimated drift for $r = 2$

Table 2: Model 2

$$\sigma^2(x) = \frac{1+x^2}{2+x^2}$$

Estimation over $[-1, 1]$.
*ris*: empirical error, *oracle*: oracle

| | | Gaussian noise | | Uniform noise | | Exponential noise | |
|---|---|---|---|---|---|---|---|
| $\Delta$ | $n$ | *risk* | *oracle* | *risk* | *oracle* | *risk* | *oracle* |
| $3.10^{-3}$ | $10^4$ | 0.018 | 2.17 | **0.0026** | 1.39 | **0.029** | 1.43 |
| $10^{-2}$ | $10^4$ | **0.0099** | 1.44 | **0.0027** | 1.37 | 0.032 | 1.29 |
| $3.10^{-2}$ | $10^4$ | 0.014 | 1.65 | **0.0028** | 1.40 | 0.038 | 1.29 |
| $10^{-1}$ | $10^4$ | 0.016 | **1.18** | 0.0065 | 1.13 | 0.035 | 1.09 |
| $3.10^{-3}$ | $10^3$ | 0.10 | 1.60 | 0.015 | 1.33 | 0.41 | 1.28 |
| $10^{-2}$ | $10^3$ | 0.14 | 1.69 | 0.015 | 1.39 | 0.38 | 1.91 |
| $3.10^{-2}$ | $10^3$ | 0.21 | 2.37 | 0.016 | **1.04** | 0.59 | 2.04 |
| $10^{-1}$ | $10^3$ | 0.14 | 1.97 | 0.019 | 1.07 | 0.32 | 1.21 |
| $10^{-2}$ | $10^2$ | 0.87 | 1.25 | 0.16 | 1.18 | 2.29 | **1.04** |
| $3.10^{-2}$ | $10^2$ | 0.96 | 1.53 | 0.11 | 1.12 | 4.94 | 1.62 |
| $10^{-1}$ | $10^2$ | 1.82 | 2.26 | 0.12 | 1.087 | 5.43 | 1.85 |

Table 3: Model 3

$$\sigma^2(x) = \frac{1+x^2}{2+x^2}$$

Estimation over $[-1, 1]$.
*ris*: empirical error, *oracle*: oracle

| | | Gaussian noise | | Uniform noise | | Laplace noise | |
|---|---|---|---|---|---|---|---|
| $\Delta$ | $n$ | *risk* | *oracle* | *risk* | *oracle* | *risk* | *oracle* |
| $3.10^{-3}$ | $10^4$ | 0.027 | 1.96 | **0.0032** | **1.11** | 0.070 | 1.68 |
| $10^{-2}$ | $10^4$ | **0.020** | 1.44 | 0.0041 | **1.10** | **0.058** | 1.48 |
| $3.10^{-2}$ | $10^4$ | 0.031 | 1.57 | 0.011 | 1.27 | 0.068 | 1.47 |
| $10^{-1}$ | $10^4$ | 0.044 | 1.23 | 0.033 | **1.10** | 0.080 | 1.41 |
| $3.10^{-3}$ | $10^3$ | 0.14 | **1.19** | 0.029 | 1.20 | 0.42 | 1.42 |
| $10^{-2}$ | $10^3$ | 0.18 | 1.89 | 0.027 | 1.56 | 0.39 | 1.26 |
| $3.10^{-2}$ | $10^3$ | 0.17 | 1.62 | 0.035 | 1.23 | 0.46 | 1.24 |
| $10^{-1}$ | $10^3$ | 0.14 | 1.40 | 0.047 | 1.28 | 0.41 | 1.81 |
| $10^{-2}$ | $10^2$ | 0.90 | 1.43 | 0.18 | **1.09** | 2.61 | **1.15** |
| $3.10^{-2}$ | $10^2$ | 1.77 | 2.07 | 0.29 | 1.29 | 5.78 | 1.86 |
| $10^{-1}$ | $10^2$ | 2.50 | 2.69 | 0.21 | **1.11** | 3.93 | 1.59 |

Table 4: Model 4

$$\sigma(x) = 1 + |x| \,.$$

Estimation over $[-1, 1]$.
*ris*: empirical error, *or*: oracle

| | | Gaussian noise | | Uniform noise | | Laplace noise | |
|---|---|---|---|---|---|---|---|
| $\Delta$ | $n$ | *risk* | *oracle* | *risk* | *oracle* | *risk* | *oracle* |
| $3.10^{-3}$ | $10^4$ | 0.23 | 2.96 | **0.07** | 1.62 | 0.25 | 2.12 |
| $10^{-2}$ | $10^4$ | **0.12** | 1.58 | **0.08** | 1.97 | **0.18** | 1.67 |
| $3.10^{-2}$ | $10^4$ | 0.22 | 1.41 | 0.13 | 1.33 | 0.42 | 1.38 |
| $10^{-1}$ | $10^4$ | 0.82 | 1.36 | 0.34 | **1.08** | 1.43 | **1.19** |
| $3.10^{-3}$ | $10^3$ | 0.42 | 1.20 | 0.22 | 1.70 | 1.05 | 1.46 |
| $10^{-2}$ | $10^3$ | 0.96 | 2.68 | 0.27 | 1.62 | 1.23 | 1.81 |
| $3.10^{-2}$ | $10^3$ | 0.79 | 1.76 | 0.67 | 1.41 | 2.53 | 2.55 |
| $10^{-1}$ | $10^3$ | 1.74 | 2.15 | 0.79 | 1.51 | 2.89 | 1.70 |
| $10^{-2}$ | $10^2$ | 1.49 | **1.18** | 0.85 | 1.27 | 4.01 | 1.06 |
| $3.10^{-2}$ | $10^2$ | 4.85 | 1.98 | 1.61 | 1.39 | 4.74 | **1.19** |
| $10^{-1}$ | $10^2$ | 4.07 | 1.78 | 10.6 | 3.78 | 12.5 | 2.06 |