



# Non-parametric estimation of the coefficients of ergodic diffusion processes based on high-frequency data

Fabienne Comte, Valentine Genon-Catalot, Yves Rozenholc

## ► To cite this version:

Fabienne Comte, Valentine Genon-Catalot, Yves Rozenholc. Non-parametric estimation of the coefficients of ergodic diffusion processes based on high-frequency data. M. Kessler, A. Lindner, M. Sorensen. Statistical Methods for Stochastic Differential Equations, Chapman & Hall/CRC Monographs on Statistics & Applied Probability, pp.341-381, 2012, SemStat series. hal-00748939

**HAL Id: hal-00748939**

**<https://hal.science/hal-00748939>**

Submitted on 6 Nov 2012

**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.

---

## CHAPTER 5

# Non-parametric estimation of the coefficients of ergodic diffusion processes based on high-frequency data

---

Fabienne Comte, Valentine Genon-Catalot and Yves Rozenholc

UFR de Mathématiques et Informatique, Université Paris Descartes – Paris 5  
45 rue des Saints-Pères, 75270 Paris cedex 06, France

### 5.1 Introduction

The content of this chapter is directly inspired by Comte, Genon-Catalot, and Rozenholc (2006; 2007). We consider non-parametric estimation of the drift and diffusion coefficients of a one-dimensional diffusion process. The main assumption on the diffusion model is that it is ergodic and geometrically  $\beta$ -mixing. The sample path is assumed to be discretely observed with a small regular sampling interval  $\Delta$ . The estimation method that we develop is based on a penalized mean square approach. This point of view is fully investigated for regression models in Comte and Rozenholc (2002, 2004). We adapt it to discretized diffusion models.

### 5.2 Model and assumptions

Let  $(X_t)_{t \geq 0}$  be a one-dimensional diffusion process with dynamics described by the stochastic differential equation:

$$dX_t = b(X_t)dt + \sigma(X_t)dW_t, \quad t \geq 0, \quad X_0 = \eta, \quad (5.1)$$

where  $(W_t)$  is a standard Brownian motion and  $\eta$  is a random variable independent of  $(W_t)$ . Consider the following assumptions:

[A1]  $-\infty \leq l < r \leq +\infty$ ,  $b$  and  $\sigma$  belong to  $C^1((l, r))$ ,  $\sigma(x) > 0$  for all  $x \in (l, r)$ .

[A2] For  $x_0, x \in (l, r)$ , let  $s(x) = \exp(-2 \int_{x_0}^x b(u)/\sigma^2(u) du)$  denote the scale density and  $m(x) = 1/[\sigma^2(x)s(x)]$  the speed density. We assume

$$\int_l^r s(x) dx = +\infty = \int_l^r s(x) dx, \quad \int_l^r m(x) dx = M < \infty.$$

[A3] The initial random variable  $\eta$  has distribution

$$\pi(x) dx = m(x)/M \mathbf{1}_{(l, r)}(x) dx.$$

Under [A1] – [A2], equation (5.1) admits a unique strong solution with state space the open interval  $(l, r)$  of the real line. Moreover, it is positive recurrent on this interval and admits as unique stationary distribution the normalized speed density  $\pi$ . With the additional assumption [A3], the process  $(X_t)$  is strictly stationary, with marginal distribution  $\pi(x) dx$ , ergodic and  $\beta$ -mixing, i.e.  $\lim_{t \rightarrow +\infty} \beta_X(t) = 0$  where  $\beta_X(t)$  denotes the  $\beta$ -mixing coefficient of  $(X_t)$ . For stationary Markov processes such as  $(X_t)$ , the  $\beta$ -mixing coefficient has the following explicit expression

$$\beta_X(t) = \int_l^r \pi(x) dx \|P_t(x, dx') - \pi(x') dx'\|_{TV}. \quad (5.2)$$

The norm  $\|\cdot\|_{TV}$  is the total variation norm and  $P_t$  denotes the transition probability (see e.g. Genon-Catalot, Jeantheau, and Larédo (2000) for a review). The statistical study relies on a stronger mixing condition which is satisfied in most standard models.

[A4] There exist constants  $K > 0$  and  $\theta > 0$  such that:

$$\beta_X(t) \leq K e^{-\theta t}. \quad (5.3)$$

In some cases, assumption [A4] can be checked directly using formula (5.2) (see Proposition 5.14 below). Otherwise, simple sufficient conditions are available (see e.g. Proposition 1 in Pardoux and Veretennikov (2001)). Lastly, we strengthen assumptions on  $b$  and  $\sigma$  to deal altogether with finite or infinite boundaries and keep a general, simple and clear framework. We also need a moment assumption for  $\pi$  and that  $\sigma$  be bounded.

[A5] Let  $I = [l, r] \cap \mathbb{R}$ .

(i) Assume that  $b \in C^1(I)$ ,  $b'$  bounded on  $I$ ,  $\sigma^2 \in C^2(I)$ ,  $(\sigma^2)''$  bounded on  $I$ .

(ii) For all  $x \in I$ ,  $\sigma^2(x) \leq \sigma_1^2$ .

[A6]  $E(\eta^8) < \infty$ .

The following property will be useful.

**Lemma 5.1** *Under Assumptions [A1]–[A3] and [A5]–[A6], for all  $t, s$  such that  $|t - s| \leq 1$ , for  $1 \leq i \leq 4$ ,  $E((X_t - X_s)^{2i}) \leq c|t - s|^i$ .*

*Proof.* Using the strict stationarity, we only need to study  $E((X_t - X_0)^{2i})$  for  $t \leq 1$ . By the Minkowski inequality,

$$(X_t - X_0)^{2i} \leq 2^{2i-1} \left[ \left( \int_0^t b(X_s) ds \right)^{2i} + \left( \int_0^t \sigma(X_s) dW_s \right)^{2i} \right].$$

For the drift term, we use the Hölder inequality, [A5] and the strict stationarity to get:

$$E \left[ \left( \int_0^t b(X_s) ds \right)^{2i} \right] \leq t^{2i-1} \int_0^t E(b^{2i}(X_s)) ds \leq t^{2i} C(1 + E(\eta^{2i})),$$

with  $C$  a constant. For the diffusion term, we use the Burkholder–Davis–Gundy inequality and obtain:

$$E \left[ \left( \int_0^t \sigma(X_s) dW_s \right)^{2i} \right] \leq C E \left( \int_0^t \sigma^2(X_s) ds \right)^i \leq C t^i \sigma_1^{2i}$$

with  $C$  a constant. This gives the result.  $\square$

### 5.3 Observations and asymptotic framework

We assume that the sample path  $X_t$  is observed at  $n + 1$  discrete instants with sampling interval  $\Delta$ . The asymptotic framework that we consider is the context of high-frequency data: the sampling interval  $\Delta = \Delta_n$  tends to 0 as  $n$  tends to infinity. Moreover, we assume that the total length  $n\Delta_n$  of the time interval where observations are taken tends to infinity. This is a classical framework for ergodic diffusion models: it allows us to estimate simultaneously the drift  $b$  and the diffusion coefficient  $\sigma$  and to enlighten us about the different rates of estimation of these two coefficients. For the penalized non-parametric method developed here, the asymptotic framework has to fulfill the following strengthened condition.

[A7] As  $n$  tends to infinity,  $\Delta = \Delta_n \rightarrow 0$  and  $n\Delta_n / \ln^2 n \rightarrow +\infty$ .

To simplify notations, we will drop the subscript and simply write  $\Delta$  for the sampling interval.

### 5.4 Estimation method

#### 5.4.1 General description

We aim at estimating functions  $b$  and  $\sigma^2$  of model (5.1) on a compact subset  $A$  of the state space  $(l, r)$ . For simplicity, and without loss of generality, we

assume from now on that

$$A = [0, 1],$$

and we set

$$b_A = b\mathbf{1}_A, \quad \sigma_A = \sigma\mathbf{1}_A. \quad (5.4)$$

The estimation method is inspired by what is done for regression models (see e.g. Comte and Rozenholc (2002, 2004)). Suppose we have observations  $(x_i, y_i)$ ,  $i = 1, \dots, n$  such that

$$y_i = f(x_i) + \text{noise}, \quad (5.5)$$

where  $f$  is unknown. We consider a regression contrast of the form

$$t \rightarrow \gamma_n(t) = \frac{1}{n} \sum_{i=1}^n (y_i - t(x_i))^2.$$

The aim is to build estimators for  $f$  by minimizing the least-square criterion  $\gamma_n(t)$ . For that purpose, we consider a collection of finite dimensional linear subspaces of  $\mathbb{L}^2([0, 1])$  and compute for each space an associated least-squares estimator. Afterwards, a data-driven procedure chooses from the resulting collection of estimators the “best” one, in a sense to be precised, through a penalization device. For adapting the method to discretized diffusion processes, we have to find a regression equation analogous to (5.5), i.e. find for  $f = b, \sigma^2$  the appropriate couple  $(x_i, y_i)$  and the adequate regression equation. Of course, starting with a diffusion model, we do not find an exact regression equation but only a regression-type equation, one for the drift and another one for the diffusion coefficient. Hence, estimators for  $b$  and for  $\sigma^2$  are built from two distinct constructions and the method does not require one to estimate the stationary density  $\pi$ .

#### 5.4.2 Spaces of approximation

Let us describe now some possible collections of spaces of approximation. We focus on two specific collections, the collection of dyadic regular piecewise polynomial spaces, denoted hereafter by [DP], and the collection of general piecewise polynomials, denoted by [GP], which is more complex. As for numerical implementation, algorithms for both collections are available and have been implemented on several examples.

In what follows, several norms for  $[0, 1]$ -supported functions are needed and the following notations will be used:

$$\begin{aligned} \|t\| &= (\int_0^1 t^2(x) dx)^{1/2}, \quad \|t\|_\pi = (\int_0^1 t^2(x) \pi(x) dx)^{1/2}, \\ \|t\|_\infty &= \sup_{x \in [0, 1]} |t(x)|. \end{aligned} \quad (5.6)$$

By our assumptions, the stationary density  $\pi$  is bounded from below and above

on every compact subset of  $(l, r)$ . Hence, let  $\pi_0, \pi_1$  denote two positive real numbers such that,

$$\forall x \in [0, 1], \quad 0 < \pi_0 \leq \pi(x) \leq \pi_1. \quad (5.7)$$

Thus, for  $[0, 1]$ -supported functions, the norms  $\|\cdot\|$  and  $\|\cdot\|_\pi$  are equivalent.

### *Dyadic regular piecewise polynomials*

Let  $r \geq 0, p \geq 0$  be integers. On each subinterval  $I_j = [(j-1)/2^p, j/2^p]$ ,  $j = 1, \dots, 2^p$ , consider  $r+1$  polynomials of degree  $\ell$ ,  $\varphi_{j,\ell}(x)$ ,  $\ell = 0, 1, \dots, r$  and set  $\varphi_{j,\ell}(x) = 0$  outside  $I_j$ . The space  $S_m$ ,  $m = (p, r)$ , is defined as generated by the  $D_m = 2^p(r+1)$  functions  $(\varphi_{j,\ell})$ . A function  $t$  in  $S_m$  may be written as

$$t(x) = \sum_{j=1}^{2^p} \sum_{\ell=0}^r t_{j,\ell} \varphi_{j,\ell}(x).$$

The collection of spaces  $(S_m, m \in \mathcal{M}_n)$  is such that, for  $r_{max}$  a fixed integer,

$$\mathcal{M}_n = \{m = (p, r), p \in \mathbb{N}, r \in \{0, 1, \dots, r_{max}\}, 2^p(r_{max} + 1) \leq N_n\}. \quad (5.8)$$

In other words,  $D_m \leq N_n$  with  $N_n \leq n$ . The maximal dimension  $N_n$  will be subject to additional constraints. The role of  $N_n$  is to bound all dimensions  $D_m$ , even when  $m$  is random. In practice, it corresponds to the maximal number of coefficients to estimate. Thus it must not be too large.

More concretely, consider the orthogonal collection in  $\mathbb{L}^2([-1, 1])$  of Legendre polynomials  $(Q_\ell, \ell \geq 0)$ , where the degree of  $Q_\ell$  is equal to  $\ell$ , generating  $\mathbb{L}^2([-1, 1])$  (see Abramowitz and Stegun (1972), p.774). They satisfy  $|Q_\ell(x)| \leq 1, \forall x \in [-1, 1]$ ,  $Q_\ell(1) = 1$  and  $\int_{-1}^1 Q_\ell^2(u) du = 2/(2\ell + 1)$ . Then we set  $P_\ell(x) = (2\ell + 1)^{1/2} Q_\ell(2x - 1)$ , to get an orthonormal basis of  $\mathbb{L}^2([0, 1])$ . Finally,

$$\varphi_{j,\ell}(x) = 2^{p/2} P_\ell(2^p x - j + 1) \mathbf{1}_{I_j}(x), \quad j = 1, \dots, 2^p, \quad \ell = 0, 1, \dots, r.$$

The space  $S_m$  has dimension  $D_m = 2^p(r+1)$  and its orthonormal basis described above satisfies

$$\left\| \sum_{j=1}^{2^p} \sum_{\ell=0}^r \varphi_{j,\ell}^2 \right\|_\infty \leq D_m(r+1).$$

Hence, using notations (5.6), for all  $t \in S_m$ ,  $\|t\|_\infty \leq (r+1)^{1/2} D_m^{1/2} \|t\|$ .

Collection [DP] is simple in the sense that one dimension  $D_m$  is associated

with a single space  $S_m$ . In particular, since  $N_n \leq n$ , the following holds:

$$\Sigma = \sum_{m \in \mathcal{M}_n} \exp(-D_m) = \sum_{r=0}^{r_{max}} \sum_{p: 2^p(r+1) \leq N_n} \exp(-2^p(r+1)) < \infty. \quad (5.9)$$

Finally, let us sum up the two key properties that are fulfilled by this collection. The collection  $(S_m)_{m \in \mathcal{M}_n}$  is composed of finite dimensional linear sub-spaces of  $\mathbb{L}^2([0, 1])$ , indexed by a set  $\mathcal{M}_n$  depending on  $n$ . The space  $S_m$  has dimension  $D_m \leq N_n \leq n$ ,  $\forall m \in \mathcal{M}_n$ , where  $N_n$  designates a maximal dimension, and  $S_m$  is equipped with an orthonormal basis  $(\varphi_\lambda)_{\lambda \in \Lambda_m}$  with  $|\Lambda_m| = D_m$ . The following holds:

( $\mathcal{H}_1$ ) Norm connection: There exists  $\Phi_0 > 0$ , such that,

$$\forall m \in \mathcal{M}_n, \forall t \in S_m, \|t\|_\infty \leq \Phi_0 D_m^{1/2} \|t\|. \quad (5.10)$$

( $\mathcal{H}_2$ ) Nesting condition: There exists a space denoted by  $\mathcal{S}_n$ , belonging to the collection, with  $\forall m \in \mathcal{M}_n, S_m \subset \mathcal{S}_n$ , with dimension denoted by  $N_n$ .

In Birgé and Massart (1998, p.337, Lemma 1), it is proved that property (5.10) is equivalent to:

$$\text{There exists } \Phi_0 > 0, \left\| \sum_{\lambda \in \Lambda_m} \varphi_\lambda^2 \right\|_\infty \leq \Phi_0^2 D_m. \quad (5.11)$$

There are other collections of spaces satisfying the above two properties and for which our proofs apply (for instance, the trigonometric spaces, or the dyadic wavelet generated spaces).

#### *General piecewise polynomials*

A more general family can be described, the collection of general piecewise polynomials spaces denoted by [GP]. We first build the largest space  $\mathcal{S}_n$  of the collection whose dimension is denoted as above by  $N_n$  ( $N_n \leq n$ ). For this, we fix an integer  $r_{max}$  and let  $d_{max}$  be an integer such that  $d_{max}(r_{max}+1) = N_n$ . The space  $\mathcal{S}_n$  is linearly spanned by piecewise polynomials of degree  $r_{max}$  on the regular subdivision of  $[0, 1]$  with step  $1/d_{max}$ . Any other space  $S_m$  of the collection is described by a multi-index  $m = (d, j_1, \dots, j_{d-1}, r_1, \dots, r_d)$  where  $d$  is the number of intervals of the partition,  $j_0 := 0 < j_1 < \dots < j_{d-1} < j_d := 1$  are integers such that  $j_i \in \{1, \dots, d_{max}-1\}$  for  $i = 1, \dots, d-1$ . The latter integers define the knots  $j_i/d_{max}$  of the subdivision. Lastly  $r_i \leq r_{max}$  is the degree of the polynomial on the interval  $[j_{i-1}/d_{max}, j_i/d_{max})$ , for  $i = 1, \dots, d$ . A function  $t$  in  $S_m$  can thus be described as

$$t(x) = \sum_{i=1}^d P_i(x) \mathbf{1}_{[j_{i-1}/d_{max}, j_i/d_{max})}(x),$$

with  $P_i$  a polynomial of degree  $r_i$ . The dimension of  $S_m$  is still denoted by  $D_m$  and is equal to  $\sum_{i=1}^d (r_i + 1)$  for all the  $\binom{d_{max}-1}{d-1}$  choices of the knots  $(j_1, \dots, j_{d-1})$ . Note that the  $P_i$ 's can still be decomposed by using the Legendre basis rescaled on the intervals  $[j_{i-1}/d_{max}, j_i/d_{max}]$ . The collection [GP] of models  $(S_m)_{m \in \mathcal{M}_n}$  is described by the set of indexes

$$\mathcal{M}_n = \{m = (d, j_1, \dots, j_{d-1}, r_1, \dots, r_d), 1 \leq d \leq d_{max}, \\ j_i \in \{1, \dots, d_{max} - 1\}, r_i \in \{0, \dots, r_{max}\}\}.$$

It is important to note that now, for all  $m \in \mathcal{M}_n$ , for all  $t \in S_m$ , since  $S_m \subset S_n$ ,

$$\|t\|_\infty \leq \sqrt{(r_{max} + 1)N_n} \|t\|. \quad (5.12)$$

Hence, the norm connection property still holds but only on the maximal space and no more on each individual space of the collection as for collection [DP]. This comes from the fact that regular partitions are involved in all spaces of [DP] and only in the maximal space of [GP]. Obviously, collection [GP] has higher complexity than [DP]. The complexity of a collection is usually evaluated through a set of weights  $(L_m)$  that must satisfy  $\sum_{m \in \mathcal{M}_n} e^{-L_m D_m} < \infty$ . For [DP],  $L_m = 1$  suits (see (5.9)). For [GP], we have to look at

$$\sum_{m \in \mathcal{M}_n} e^{-L_m D_m} = \sum_{d=1}^{d_{max}} \sum_{1 \leq j_1 < \dots < j_{d-1} < d_{max}} \sum_{0 \leq r_1, \dots, r_d \leq r_{max}} e^{-L_m \sum_{i=1}^d (r_i + 1)}.$$

From the equality above, we deduce that the choice

$$L_m D_m = D_m + \ln \left( \binom{d_{max} - 1}{d - 1} \right) + d \ln(r_{max} + 1) \quad (5.13)$$

can suit. Actually, this relation guides the choice of the penalty function used in the practical implementation. To see more clearly what orders of magnitude are involved, let us choose  $L_m = L_n$  for all  $m \in \mathcal{M}_n$ . Then, we have a further bound for the series:

$$\begin{aligned} \sum_{m \in \mathcal{M}_n} e^{-L_m D_m} &\leq \sum_{d=1}^{d_{max}} \left( \binom{d_{max} - 1}{d - 1} \right) (r_{max} + 1)^d e^{-d L_n} \\ &\leq \sum_{d=0}^{d_{max}-1} \left( \binom{d_{max} - 1}{d} \right) [(r_{max} + 1) e^{-L_n}]^{d+1} \\ &\leq (r_{max} + 1) [1 + (r_{max} + 1) e^{-L_n}]^{d_{max}-1} \\ &\leq (r_{max} + 1) \exp(d_{max} (r_{max} + 1) e^{-L_n}) \\ &\leq (r_{max} + 1) \exp(N_n e^{-L_n}). \end{aligned}$$



Thus  $L_m = L_n = \ln(N_n)$  ensures that the series is bounded. (For more details on these collections, see *e.g.* Comte and Rozenholc (2004) or Baraud, Comte, and Viennet (2001b).

## 5.5 Drift estimation

### 5.5.1 Drift estimators: Statements of the results

The regression-type equation for the drift is as follows:

$$Y_{k\Delta} := \frac{X_{(k+1)\Delta} - X_{k\Delta}}{\Delta} = b(X_{k\Delta}) + Z_{k\Delta} + R_{k\Delta} \quad (5.14)$$

where

$$Z_{k\Delta} = \frac{1}{\Delta} \int_{k\Delta}^{(k+1)\Delta} \sigma(X_s) dW_s, \quad R_{k\Delta} = \frac{1}{\Delta} \int_{k\Delta}^{(k+1)\Delta} (b(X_s) - b(X_{k\Delta})) ds. \quad (5.15)$$

The couple  $(X_{k\Delta}, Y_{k\Delta})$  stands for the data  $(x_k, y_k)$ . The term  $Z_{k\Delta}$  is a martingale increment (with respect to the filtration  $\mathcal{F}_{k\Delta} = \sigma(X_s, s \leq k\Delta)$ ) and plays the role of the noise term. The term  $R_{k\Delta}$  is a remainder due to the discretization. Now, we consider a collection which may be either [DP] or [GP]. For  $S_m$  a space of the collection and for  $t \in S_m$ , we set

$$\gamma_n(t) = \frac{1}{n} \sum_{k=1}^n [Y_{k\Delta} - t(X_{k\Delta})]^2. \quad (5.16)$$

We define an estimator  $\hat{b}_m$  of  $b_A$  belonging to  $S_m$  as any solution of:

$$\hat{b}_m = \arg \min_{t \in S_m} \gamma_n(t). \quad (5.17)$$

Note that, with this definition, only the random  $\mathbb{R}^n$ -vector  $(\hat{b}_m(X_\Delta), \dots, \hat{b}_m(X_{n\Delta}))'$  is uniquely defined. Indeed, let  $\Pi_m$  denote the orthogonal projection (with respect to the inner product of  $\mathbb{R}^n$ ) onto the subspace  $\{(t(X_\Delta), \dots, t(X_{n\Delta}))', t \in S_m\}$  of  $\mathbb{R}^n$ . Then  $(\hat{b}_m(X_\Delta), \dots, \hat{b}_m(X_{n\Delta}))' = \Pi_m Y$  where  $Y = (Y_\Delta, \dots, Y_{n\Delta})'$ . Any function  $t$  in  $S_m$  such that  $t(X_{k\Delta}) = \hat{b}_m(X_{k\Delta})$ ,  $k = 1, \dots, n$ , is a solution of (5.17).

This is the reason why we adopt a specific definition of the risk of an estimator. Consider the following empirical norm for a function  $t$ :

$$\|t\|_n^2 = \frac{1}{n} \sum_{k=1}^n t^2(X_{k\Delta}). \quad (5.18)$$

The risk of  $\hat{b}_m$  is defined as the expectation of the empirical norm:

$$\mathbb{E}(\|\hat{b}_m - b_A\|_n^2).$$

Note that, for a deterministic function  $t$ ,  $\mathbb{E}(\|t\|_n^2) = \|t\|_\pi^2 = \int t^2(x)\pi(x)dx$ .

The following proposition provides an upper bound for the risk of an estimator  $\hat{b}_m$  with fixed  $m$ . Let  $b_m$  denote the orthogonal projection of  $b$  on  $S_m$ .

**Proposition 5.2** *Assume that [A1] – [A7] hold. Consider a space  $S_m$  in collection [DP] or [GP] with maximal dimension satisfying  $N_n = o(n\Delta/\ln^2(n))$ . Then the estimator  $\hat{b}_m$  of  $b$  is such that (see (5.4))*

$$\mathbb{E}(\|\hat{b}_m - b_A\|_n^2) \leq 7\pi_1 \|b_m - b_A\|^2 + K \frac{\sigma_1^2 D_m}{n\Delta} + K' \Delta + \frac{K''}{n\Delta}, \quad (5.19)$$

where  $K, K'$  and  $K''$  are some positive constants.

As usual, there appears to be a squared bias term  $\|b_m - b_A\|^2$  and a variance term of order  $D_m/(n\Delta)$ , plus additional terms due to the discretization. It is standard for diffusion models in high-frequency data to assume that  $\Delta = o(1/(n\Delta))$  so that the two last terms in (5.19) are negligible with respect to the variance term. It remains to select the dimension  $D_m$  that leads to the best compromise between the squared bias term and the variance term.

Consider the case [DP]. To compare the result of Proposition 5.2 with the optimal non-parametric rates exhibited by Hoffmann (1999), let us assume that  $b_A$  belongs to a ball of some Besov space,  $b_A \in \mathcal{B}_{\alpha,2,\infty}([0,1])$ , and that  $r+1 \geq \alpha$ . Then, for  $\|b_A\|_{\alpha,2,\infty} \leq L$ , it is known that  $\|b_A - b_m\|^2 \leq C(\alpha, L) D_m^{-2\alpha}$  (see DeVore and Lorentz (1993, p.359) or Lemma 12 in Barron, Birgé, and Massart (1999)). Thus, choosing  $D_m = (n\Delta)^{1/(2\alpha+1)}$ , we obtain

$$\mathbb{E}(\|\hat{b}_m - b_A\|_n^2) \leq C(\alpha, L)(n\Delta)^{-2\alpha/(2\alpha+1)} + K' \Delta + \frac{K''}{n\Delta}.$$

The first term  $(n\Delta)^{-2\alpha/(2\alpha+1)}$  is exactly the optimal non-parametric rate (see Hoffmann (1999)). Under the condition  $\Delta = o(1/(n\Delta))$ , the last two terms in (5.19) are negligible with respect to  $(n\Delta)^{-2\alpha/(2\alpha+1)}$ .

As a second step, we must ensure an automatic selection of  $D_m$ , which does not use any knowledge on  $b$ , and in particular which does not require knowing its regularity  $\alpha$ . This selection is done by defining

$$\hat{m} = \arg \min_{m \in \mathcal{M}_n} [\gamma_n(\hat{b}_m) + \text{pen}(m)], \quad (5.20)$$

with  $\text{pen}(m)$  a penalty to be properly chosen. We denote by  $\hat{b}_{\hat{m}}$  the resulting estimator and we need to determine  $\text{pen}(\cdot)$  such that, ideally,

$$\mathbb{E}(\|\hat{b}_{\hat{m}} - b_A\|_n^2) \leq C \inf_{m \in \mathcal{M}_n} \left( \|b_A - b_m\|^2 + \frac{\sigma_1^2 D_m}{n\Delta} \right) + K' \Delta + \frac{K''}{n\Delta},$$

with  $C$  a constant, which should not be too large. We almost reach this aim.

**Theorem 5.3** Assume that [A1] – [A7] hold and consider the nested collection of models [DP] with  $L_m = 1$  or the collection [GP] with  $L_m$  given by (5.13), both with maximal dimension  $N_n = o(n\Delta/\ln^2(n))$ . Let

$$\text{pen}(m) \geq \kappa \sigma_1^2 \frac{(1 + L_m) D_m}{n\Delta}, \quad (5.21)$$

where  $\kappa$  is a universal constant. Then the estimator  $\hat{b}_{\hat{m}}$  of  $b$  with  $\hat{m}$  defined in (5.20) is such that

$$\mathbb{E}(\|\hat{b}_{\hat{m}} - b_A\|_n^2) \leq C \inf_{m \in \mathcal{M}_n} (\|b_m - b_A\|^2 + \text{pen}(m)) + K' \Delta + \frac{K''}{n\Delta}. \quad (5.22)$$

Inequality (5.22) shows that the adaptive estimator automatically realizes the bias-variance compromise. Nevertheless, some comments need to be made. It is possible to choose the equality in (5.21) but this is not what is done in practice. It is better to have additional terms to avoid underpenalization. The constant  $\kappa$  in the penalty is numerical and must be calibrated for the problem. Another important point is that  $\sigma_1^2$  is unknown. In practice, it is replaced by a rough estimator.

### 5.5.2 Proof of Proposition 5.2

Let us set (see (5.15)), for any function  $t(\cdot)$ ,

$$\nu_n(t) = \frac{1}{n} \sum_{k=1}^n t(X_{k\Delta}) Z_{k\Delta}, \quad R_n(t) = \frac{1}{n} \sum_{k=1}^n t(X_{k\Delta}) R_{k\Delta}. \quad (5.23)$$

Using (5.14) – (5.16) – (5.18), we have:

$$\gamma_n(t) - \gamma_n(b) = \|t - b\|_n^2 + 2\nu_n(b - t) + 2R_n(b - t).$$

Recall that  $b_m$  denotes the orthogonal projection of  $b$  on  $S_m$ . By definition of  $\hat{b}_m$ ,  $\gamma_n(\hat{b}_m) \leq \gamma_n(b_m)$ . So,  $\gamma_n(\hat{b}_m) - \gamma_n(b) \leq \gamma_n(b_m) - \gamma_n(b)$ . This implies

$$\|\hat{b}_m - b\|_n^2 \leq \|b_m - b\|_n^2 + 2\nu_n(\hat{b}_m - b_m) + 2R_n(\hat{b}_m - b_m).$$

The functions  $\hat{b}_m$  and  $b_m$  being  $A$ -supported, we can cancel the terms  $\|b \mathbf{1}_{A^c}\|_n^2$  that appear in both sides of the inequality. This yields

$$\|\hat{b}_m - b_A\|_n^2 \leq \|b_m - b_A\|_n^2 + 2\nu_n(\hat{b}_m - b_m) + 2R_n(\hat{b}_m - b_m).$$

Recall notations (5.6) and let

$$B_m^\pi = \{t \in S_m, \|t\|_\pi = 1\}.$$

We use the standard inequality  $2xy \leq \theta^2 x^2 + y^2/\theta^2$  which holds for all  $\theta \neq 0$  with  $\theta^2 = 8$ :

$$2\nu_n(\hat{b}_m - b_m) \leq 2\|\hat{b}_m - b_m\|_\pi \sup_{t \in B_m^\pi} |\nu_n(t)| \leq \frac{1}{8} \|\hat{b}_m - b_m\|_\pi^2 + 8 \sup_{t \in B_m^\pi} [\nu_n(t)]^2.$$

Similarly,

$$\begin{aligned} 2R_n(\hat{b}_m - b_m) &\leq 2\|\hat{b}_m - b_m\|_n \left( \frac{1}{n} \sum_{k=1}^n R_{k\Delta}^2 \right)^{1/2} \\ &\leq \frac{1}{8} \|\hat{b}_m - b_m\|_n^2 + 8 \frac{1}{n} \sum_{k=1}^n R_{k\Delta}^2. \end{aligned}$$

Because the  $\mathbb{L}_\pi^2$ -norm,  $\|\cdot\|_\pi$ , and the empirical norm (5.18) are not equivalent, we must introduce a set on which they are, and afterwards prove that this set has small probability. Let us define (see (5.8))

$$\Omega_n = \left\{ \omega / \left| \frac{\|t\|_n^2}{\|t\|_\pi^2} - 1 \right| \leq \frac{1}{2}, \quad \forall t \in \bigcup_{m, m' \in \mathcal{M}_n} (S_m + S_{m'}) / \{0\} \right\}. \quad (5.24)$$

On  $\Omega_n$ ,  $\|\hat{b}_m - b_m\|_\pi^2 \leq 2\|\hat{b}_m - b_m\|_n^2$ , and  $\|\hat{b}_m - b_m\|_n^2 \leq 2(\|\hat{b}_m - b_A\|_n^2 + \|b_m - b_A\|_n^2)$ . Hence, some elementary computations yield:

$$\frac{1}{4} \|\hat{b}_m - b_A\|_n^2 \mathbf{1}_{\Omega_n} \leq \frac{7}{4} \|b_m - b_A\|_n^2 + 8 \sup_{t \in B_m^\pi} [\nu_n(t)]^2 + \frac{8}{n} \sum_{k=1}^n R_{k\Delta}^2.$$

Now, using [A5] and Lemma 5.1, we get

$$\mathbb{E}(R_{k\Delta}^2) \leq \frac{1}{\Delta} \mathbb{E} \int_{k\Delta}^{(k+1)\Delta} (b(X_s) - b(X_{k\Delta}))^2 ds \leq c' \Delta.$$

Consequently,

$$\mathbb{E}(\|\hat{b}_m - b_A\|_n^2 \mathbf{1}_{\Omega_n}) \leq 7\|b_m - b_A\|_\pi^2 + 32 \mathbb{E} \left( \sup_{t \in B_m^\pi} [\nu_n(t)]^2 \right) + 32c' \Delta. \quad (5.25)$$

Consider a basis of  $S_m$ , say  $\{\psi_\lambda, \lambda \in J_m\}$ , which is orthonormal with respect to  $L_\pi^2$ , and with  $|J_m| = D_m$ . For  $t \in B_m^\pi$ ,

$$[\nu_n(t)]^2 \leq \sum_{\lambda \in J_m} [\nu_n^2(\psi_\lambda)].$$

Using the martingale property of (5.23) and the bound of  $\sigma^2(\cdot)$ , we get:

$$\begin{aligned} \mathbb{E}[\nu_n^2(\psi_\lambda)] &= \frac{1}{n^2 \Delta^2} \sum_{k=1}^n \mathbb{E} \left\{ \psi_\lambda^2(X_{k\Delta}) \int_{k\Delta}^{(k+1)\Delta} \sigma^2(X_s) ds \right\} \\ &\leq \frac{\sigma_1^2}{n\Delta} \int \psi_\lambda^2(x) \pi(x) dx = \frac{\sigma_1^2}{n\Delta}. \end{aligned}$$

Therefore,

$$\mathbb{E} \left( \sup_{t \in B_m^\pi} [\nu_n(t)]^2 \right) \leq \frac{\sigma_1^2}{n\Delta} D_m.$$

Gathering bounds, and using the upper bound  $\pi_1$  defined in (5.7), we get

$$\mathbb{E}(\|\hat{b}_m - b_A\|_n^2 \mathbf{1}_{\Omega_n^c}) \leq 7\pi_1 \|b_m - b_A\|^2 + 32 \frac{\sigma_1^2 D_m}{n\Delta} + 32c'\Delta.$$

Now, it remains to deal with  $\Omega_n^c$ . Since  $\|\hat{b}_m - b_A\|_n^2 \leq \|\hat{b}_m - b\|_n^2$ , it is enough to check that  $\mathbb{E}(\|\hat{b}_m - b\|_n^2 \mathbf{1}_{\Omega_n^c}) \leq c/n$ . Write  $Y_{k\Delta} = b(X_{k\Delta}) + \varepsilon_{k\Delta}$  with  $\varepsilon_{k\Delta} = Z_{k\Delta} + R_{k\Delta}$ . Recall that  $\Pi_m$  denotes the orthogonal projection (with respect to the inner product of  $\mathbb{R}^n$ ) onto the subspace  $\{(t(X_\Delta), \dots, t(X_{n\Delta}))', t \in S_m\}$  of  $\mathbb{R}^n$ . We have  $(\hat{b}_m(X_\Delta), \dots, \hat{b}_m(X_{n\Delta}))' = \Pi_m Y$  where  $Y = (Y_\Delta, \dots, Y_{n\Delta})'$ . Using the same notation for the function  $t$  and the vector  $(t(X_\Delta), \dots, t(X_{n\Delta}))'$ , we see that

$$\|b - \hat{b}_m\|_n^2 = \|b - \Pi_m b\|_n^2 + \|\Pi_m \varepsilon\|_n^2 \leq \|b\|_n^2 + n^{-1} \sum_{k=1}^n \varepsilon_{k\Delta}^2.$$

Therefore,

$$\begin{aligned} \mathbb{E}(\|b - \hat{b}_m\|_n^2 \mathbf{1}_{\Omega_n^c}) &\leq \mathbb{E}(\|b\|_n^2 \mathbf{1}_{\Omega_n^c}) + \frac{1}{n} \sum_{k=1}^n \mathbb{E}(\varepsilon_{k\Delta}^2 \mathbf{1}_{\Omega_n^c}) \\ &\leq \left( \mathbb{E}^{1/2}(b^4(X_0)) + \mathbb{E}^{1/2}(\varepsilon_\Delta^4) \right) \mathbb{P}^{1/2}(\Omega_n^c). \end{aligned}$$

Using [A5], we have  $\mathbb{E}(b^4(X_0)) \leq c(1 + \mathbb{E}(X_0^4)) = K$ . With the Burkholder-Davis-Gundy inequality, we find

$$\mathbb{E}(\varepsilon_\Delta^4) \leq 2^3 \left\{ \frac{1}{\Delta} \int_0^\Delta \mathbb{E}[(b(X_s) - b(X_\Delta))^4] ds + \frac{36}{\Delta^3} \mathbb{E} \left( \int_0^\Delta \sigma^4(X_s) ds \right) \right\}.$$

Under [A1] – [A3], [A5] – [A6] and using Lemma 5.1, we obtain  $\mathbb{E}(\varepsilon_\Delta^4) \leq C(1 + \sigma_1^4/\Delta^2) := C'/\Delta^2$ . The next lemma enables us to complete the proof.

**Lemma 5.4** *Let  $\Omega_n$  be defined by (5.24). Then, if  $N_n \leq O(n\Delta_n/\ln^2(n))$*

$$\mathbb{P}(\Omega_n^c) \leq \frac{c}{n^4}. \quad (5.26)$$

The proof of this technical lemma is given in Comte et al. (2007) and relies on inequalities proved in Baraud, Comte, and Viennet (2001a). We stress the fact that it is for this lemma that we need the exponential  $\beta$ -mixing assumption for  $(X_t)$ . It is also for this lemma that we have constraints on the maximal dimension  $N_n$ .

Now, we gather all terms and use (5.26) to get (5.19).  $\square$

## 5.5.3 Proof of Theorem 5.3

The proof relies on the following Bernstein-type inequality:

**Lemma 5.5** *Under the assumptions of Theorem 5.3, for any positive numbers  $\epsilon$  and  $v$ , we have (see (5.23)):*

$$\mathbb{P}(\nu_n(t) \geq \epsilon, \|t\|_n^2 \leq v^2) \leq \exp\left(-\frac{n\Delta\epsilon^2}{2\sigma_1^2 v^2}\right).$$

*Proof of Lemma 5.5.* Consider the process:

$$H_u^n = H_u = \sum_{k=1}^n \mathbf{1}_{[k\Delta, (k+1)\Delta)}(u) t(X_{k\Delta}) \sigma(X_u)$$

which satisfies  $H_u^2 \leq \sigma_1^2 \|t\|_\infty^2$  for all  $u \geq 0$ . Then, denoting by  $M_s = \int_0^s H_u dW_u$ , we get that

$$M_{(n+1)\Delta} = \sum_{k=1}^n t(X_{k\Delta}) \int_{k\Delta}^{(k+1)\Delta} \sigma(X_s) dW_s = n\Delta\nu_n(t)$$

and that

$$\langle M \rangle_{(n+1)\Delta} = \sum_{k=1}^n t^2(X_{k\Delta}) \int_{k\Delta}^{(k+1)\Delta} \sigma^2(X_s) ds.$$

Moreover,  $\langle M \rangle_s = \int_0^s H_u^2 du \leq n\sigma_1^2 \Delta \|t\|_n^2$ ,  $\forall s \geq 0$ , so that  $(M_s)$  and  $\exp(\lambda M_s - \lambda^2 \langle M \rangle_s / 2)$  are martingales with respect to the filtration  $\mathcal{F}_s = \sigma(X_u, u \leq s)$ . Therefore, for all  $s \geq 0$ ,  $c > 0$ ,  $d > 0$ ,  $\lambda > 0$ ,

$$\mathbb{P}(M_s \geq c, \langle M \rangle_s \leq d) \leq \mathbb{P}\left(e^{\lambda M_s - \frac{\lambda^2}{2} \langle M \rangle_s} \geq e^{\lambda c - \frac{\lambda^2}{2} d}\right) \leq e^{-(\lambda c - \frac{\lambda^2}{2} d)}.$$

Therefore,

$$\mathbb{P}(M_s \geq c, \langle M \rangle_s \leq d) \leq \inf_{\lambda > 0} e^{-(\lambda c - \frac{\lambda^2}{2} d)} = e^{-\frac{c^2}{2d}}.$$

Finally,

$$\begin{aligned} \mathbb{P}(\nu_n(t) \geq \epsilon, \|t\|_n^2 \leq v^2) &= \mathbb{P}(M_{(n+1)\Delta} \geq n\Delta\epsilon, \langle M \rangle_{(n+1)\Delta} \leq nv^2\sigma_1^2\Delta) \\ &\leq \exp\left(-\frac{(n\Delta\epsilon)^2}{2nv^2\sigma_1^2\Delta}\right) = \exp\left(-\frac{n\epsilon^2\Delta}{2v^2\sigma_1^2}\right). \quad \square \end{aligned}$$

Now we turn to the proof of Theorem 5.3. As in the proof of Proposition 5.2, we have to split  $\|\hat{b}_{\hat{m}} - b_A\|_n^2 = \|\hat{b}_{\hat{m}} - b_A\|_n^2 \mathbf{1}_{\Omega_n} + \|\hat{b}_{\hat{m}} - b_A\|_n^2 \mathbf{1}_{\Omega_n^c}$ . For the study on  $\Omega_n^c$ , the end of the proof of Proposition 5.2 can be used.

It remains to look at what happens on  $\Omega_n$ . Let us introduce the notation

$$G_m(m') = \sup_{t \in S_m + S_{m'}, \|t\|_\pi = 1} |\nu_n(t)|.$$

From the definition of  $\hat{b}_{\hat{m}}$ , we have,  $\forall m \in \mathcal{M}_n$ ,  $\gamma_n(\hat{b}_{\hat{m}}) + \text{pen}(\hat{m}) \leq \gamma_n(b_m) + \text{pen}(m)$ . We proceed as in the proof of Proposition 5.2 with the additional penalty terms (see (5.25)) and obtain

$$\begin{aligned} \mathbb{E}(\|\hat{b}_{\hat{m}} - b_A\|_n^2 \mathbf{1}_{\Omega_n}) &\leq 7\pi_1 \|b_m - b_A\|^2 + 4\text{pen}(m) \\ &\quad + 32\mathbb{E}([G_m(\hat{m})]^2 \mathbf{1}_{\Omega_n}) - 4\mathbb{E}(\text{pen}(\hat{m})) + 32c'\Delta. \end{aligned}$$

The main problem here is to control the supremum of  $\nu_n(t)$  on a random ball (which depends on the random  $\hat{m}$ ). This will be done using Lemma 5.5 and Proposition 5.6 below, proceeding first as follows. We plug in a function  $p(m, m')$ , which will in turn fix the penalty:

$$\begin{aligned} [G_m(\hat{m})]^2 \mathbf{1}_{\Omega_n} &\leq \{([G_m(\hat{m})]^2 - p(m, \hat{m})) \mathbf{1}_{\Omega_n}\}_+ + p(m, \hat{m}) \\ &\leq \sum_{m' \in \mathcal{M}_n} \{([G_m(m')]^2 - p(m, m')) \mathbf{1}_{\Omega_n}\}_+ + p(m, \hat{m}). \end{aligned}$$

The penalty  $\text{pen}(\cdot)$  is chosen such that  $8p(m, m') \leq \text{pen}(m) + \text{pen}(m')$ . More precisely, the next proposition determines the choice of  $p(m, m')$ .

**Proposition 5.6** *Under the assumptions of Theorem 5.3, there exists a numerical constant  $\kappa_1$  such that, for  $p(m, m') = \kappa_1 \sigma_1^2 (D_m + (1 + L_{m'})D_{m'})/(n\Delta)$ , we have*

$$\mathbb{E}\{([G_m(m')]^2 - p(m, m')) \mathbf{1}_{\Omega_n}\}_+ \leq c\sigma_1^2 \frac{e^{-L_{m'}D_{m'}}}{n\Delta}.$$

*Proof of Proposition 5.6.* The result of Proposition 5.6 follows from Lemma 5.5 applying the  $\mathbb{L}_\pi^2$ -chaining technique used in Baraud et al. (2001b) (see Proposition 6.1, p.42, and Section 7, pp. 44–47, Lemma 7.1, with  $s^2 = \sigma_1^2/\Delta$ ).  $\square$

Recall that the weights  $L_m$  are such that  $\Sigma = \sum_{m' \in \mathcal{M}_n} e^{-L_{m'}D_{m'}} < \infty$ . Thus, the result of Theorem 5.3 follows from Proposition 5.6 with  $\text{pen}(m) \geq \kappa\sigma_1^2(1 + L_m)D_m/(n\Delta)$  and  $\kappa = 8\kappa_1$ .  $\square$

#### 5.5.4 Bound for the $\mathbb{L}^2$ -risk

In Theorem 5.3, the risk of  $\hat{b}_{\hat{m}}$  is not measured as a standard  $\mathbb{L}^2$ -risk. In this paragraph, we prove that a simple truncation of  $\hat{b}_{\hat{m}}$  allows to study an integrated loss over a compact set instead of our empirical loss.

**Proposition 5.7** *Let*

$$\tilde{b}^* = \begin{cases} \tilde{b} & \text{if } \|\tilde{b}\| \leq k_n \\ 0 & \text{else,} \end{cases}$$

where  $\tilde{b} = \hat{b}_{\hat{m}}$  and  $k_n = O(n)$ . Then,

$$\mathbb{E}(\|\tilde{b}^* - b_A\|^2) \leq C \inf_{m \in \mathcal{M}_n} (\|b_m - b_A\|^2 + \text{pen}(m)) + K'\Delta + \frac{K''}{n\Delta}.$$

*Proof.* Recall that  $\|\tilde{b}^* - b_A\|^2 \leq (1/\pi_0)\|\tilde{b}^* - b_A\|_\pi^2$ . Then, we decompose the  $\mathbb{L}_\pi^2$ -norm into:

$$\begin{aligned} \|\tilde{b}^* - b_A\|_\pi^2 &= \|\tilde{b}^* - b_A\|_\pi^2 \mathbf{1}_{\|\tilde{b}\| \leq k_n} \mathbf{1}_{\Omega_n} + \|\tilde{b}^* - b_A\|_\pi^2 \mathbf{1}_{\|\tilde{b}\| > k_n} \mathbf{1}_{\Omega_n} + \|\tilde{b}^* - b_A\|_\pi^2 \mathbf{1}_{\Omega_n^c} \\ &= T_1 + T_2 + T_3. \end{aligned}$$

First, it is easy to see that

$$\mathbb{E}(T_3) \leq 2(\pi_1 k_n^2 + \|b_A\|_\pi^2) \mathbb{P}(\Omega_n^c),$$

and with  $k_n = O(n)$ , as we know that  $\mathbb{P}(\Omega_n^c) \leq c/n^4$ , we get a negligible term of order  $1/n^2$ .

Next,  $T_1$  can be studied as above, except that some constants are increased:

$$\begin{aligned} T_1 \leq \|\tilde{b}^* - b_A\|_\pi^2 \mathbf{1}_{\Omega_n} &\leq 2(\|\tilde{b} - b_m\|_\pi^2 + \|b_m - b_A\|_\pi^2) \mathbf{1}_{\Omega_n} \\ &\leq 4\|\tilde{b} - b_m\|_n^2 \mathbf{1}_{\Omega_n} + 2\|b_m - b_A\|_\pi^2 \mathbf{1}_{\Omega_n} \\ &\leq 8\|\tilde{b} - b_A\|_n^2 \mathbf{1}_{\Omega_n} + 8\|b_m - b_A\|_n^2 \mathbf{1}_{\Omega_n} \\ &\quad + 2\|b_m - b_A\|_\pi^2 \mathbf{1}_{\Omega_n} \end{aligned}$$

and we can use the bound obtained in Theorem 5.3 to get that, for all  $m \in \mathcal{M}_n$ ,

$$\mathbb{E}(T_1) \leq C(\|b_m - b_A\|^2 + \text{pen}(m)) + K\Delta + \frac{K'}{n\Delta}.$$

Lastly,  $T_2 = \|b_A\|_\pi^2 \mathbf{1}_{\|\tilde{b}\| > k_n} \mathbf{1}_{\Omega_n}$ . On  $\Omega_n$ ,

$$\|\tilde{b}\|^2 \leq \frac{1}{\pi_0} \|\tilde{b}\|_\pi^2 \leq \frac{3}{2\pi_0} \|\tilde{b}\|_n^2 \leq \frac{3}{\pi_0} (\|b_A - \tilde{b}\|_n^2 + \|b_A\|_n^2)$$

and the study of this term leads to the bound

$$\|b_A - \tilde{b}\|_n^2 \leq \|b_A\|_n^2 + \frac{1}{n} \sum_{k=1}^n \varepsilon_{k\Delta}^2$$

with  $\mathbb{E}(\varepsilon_{k\Delta}^2) \leq c/\Delta$ . It follows that, with Markov's inequality,

$$\begin{aligned} \mathbb{E}(T_2) &\leq \|b_A\|_\pi^2 \mathbb{P}(\{\|\tilde{b}\| \geq k_n\} \cap \Omega_n) \\ &\leq \|b_A\|_\pi^2 \left( \mathbb{P}(6\|b\|_n^2 \geq \pi_0 k_n^2) + \mathbb{P}\left(\frac{3}{n} \sum_{k=1}^n \varepsilon_{k\Delta}^2 \geq \pi_0 k_n^2\right) \right) \\ &\leq \|b_A\|_\pi^2 \left( \frac{4\|b_A\|_\pi^2}{\pi_0 k_n^2} + \frac{c}{\Delta} \frac{3}{\pi_0 k_n^2} \right) = o\left(\frac{1}{n}\right), \end{aligned}$$

since  $k_n = O(n)$ .

Gathering all terms gives that Theorem 5.3 extends to  $\mathbb{E}(\|\tilde{b}^* - b_A\|^2)$ .  $\square$



### 5.6 Diffusion coefficient estimation

#### 5.6.1 Diffusion coefficient estimator: Statement of the results

For diffusion coefficient estimation under our asymptotic framework, it is now well known that rates of convergence are faster than for drift estimation. This is the reason why the regression-type equation has to be more precise than for  $b$ . We set

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

The couple of data is now  $(U_{k\Delta}, X_{k\Delta})$ . The regression-type equation is as follows:

$$U_{k\Delta} = \sigma^2(X_{k\Delta}) + V_{k\Delta} + \tau_{k\Delta}, \quad (5.27)$$

where  $V_{k\Delta} = V_{k\Delta}^{(1)} + V_{k\Delta}^{(2)} + V_{k\Delta}^{(3)}$  with

$$V_{k\Delta}^{(1)} = \frac{1}{\Delta} \left[ \left\{ \int_{k\Delta}^{(k+1)\Delta} \sigma(X_s) dW_s \right\}^2 - \int_{k\Delta}^{(k+1)\Delta} \sigma^2(X_s) ds \right],$$

$$V_{k\Delta}^{(2)} = \frac{1}{\Delta} \int_{k\Delta}^{(k+1)\Delta} ((k+1)\Delta - s)(\sigma^2)'(X_s)\sigma(X_s) dW_s,$$

$$V_{k\Delta}^{(3)} = 2b(X_{k\Delta}) \int_{k\Delta}^{(k+1)\Delta} \sigma(X_s) dW_s,$$

$\tau_{k\Delta} = \tau_{k\Delta}^{(1)} + \tau_{k\Delta}^{(2)} + \tau_{k\Delta}^{(3)}$  with

$$\tau_{k\Delta}^{(1)} = \frac{1}{\Delta} \left( \int_{k\Delta}^{(k+1)\Delta} b(X_s) ds \right)^2,$$

$$\tau_{k\Delta}^{(2)} = \frac{2}{\Delta} \int_{k\Delta}^{(k+1)\Delta} (b(X_s) - b(X_{k\Delta})) ds \int_{k\Delta}^{(k+1)\Delta} \sigma(X_s) dW_s,$$

$$\tau_{k\Delta}^{(3)} = \frac{1}{\Delta} \int_{k\Delta}^{(k+1)\Delta} [(k+1)\Delta - s]\psi(X_s) ds,$$

and

$$\psi = \frac{\sigma^2}{2}(\sigma^2)'' + b(\sigma^2)' = L\sigma^2, \quad (5.28)$$

where  $Lf = \frac{\sigma^2}{2}f'' + bf'$  is the infinitesimal generator of (5.1). The above relations are obtained by applying the Itô and the Fubini formulae. The term  $V_{k\Delta}$  is a sum of martingale increments whose variances have different orders. The term  $V_{k\Delta}^{(1)}$  plays the role of the main noise. The term  $\tau_{k\Delta}$  is a remainder due to the discretization and to the presence of the drift. The scheme is similar

to what is done for the drift. To estimate  $\sigma^2$  on  $A = [0, 1]$ , we define

$$\hat{\sigma}_m^2 = \arg \min_{t \in S_m} \check{\gamma}_n(t), \text{ with } \check{\gamma}_n(t) = \frac{1}{n} \sum_{k=1}^n [U_{k\Delta} - t(X_{k\Delta})]^2. \quad (5.29)$$

And, we obtain the following result.

**Proposition 5.8** *Assume that [A1]-[A7] hold and consider a model  $S_m$  in collection [DP] or [GP] with maximal dimension  $N_n = o(n\Delta/\ln^2(n))$ . Then the estimator  $\hat{\sigma}_m^2$  of  $\sigma^2$  defined by (5.29) is such that*

$$\mathbb{E}(\|\hat{\sigma}_m^2 - \sigma_A^2\|_n^2) \leq 7\pi_1 \|\sigma_m^2 - \sigma_A^2\|^2 + K \frac{\sigma_1^4 D_m}{n} + K' \Delta^2 + \frac{K''}{n}, \quad (5.30)$$

where  $K, K', K''$  are some positive constants.

Let us make some comments on the rates of convergence for estimators built with [DP]. If  $\sigma_A^2$  belongs to a ball of some Besov space, say  $\sigma_A^2 \in \mathcal{B}_{\alpha, 2, \infty}([0, 1])$ , and  $\|\sigma_A^2\|_{\alpha, 2, \infty} \leq L$ , with  $r + 1 \geq \alpha$ , then  $\|\sigma_A^2 - \sigma_m^2\|^2 \leq C(\alpha, L) D_m^{-2\alpha}$ . Therefore, if we choose  $D_m = n^{1/(2\alpha+1)}$ , we obtain

$$\mathbb{E}(\|\hat{\sigma}_m^2 - \sigma_A^2\|_n^2) \leq C(\alpha, L) n^{-2\alpha/(2\alpha+1)} + K' \Delta^2 + \frac{K''}{n}.$$

The first term  $n^{-2\alpha/(2\alpha+1)}$  is the optimal non-parametric rate proved by Hoffmann (1999). Moreover, under the standard condition  $\Delta^2 = o(1/n)$ , the last two terms are  $O(1/n)$ , i.e. negligible with respect to  $n^{-2\alpha/(2\alpha+1)}$ .

As previously, the second step is to ensure an automatic selection of  $D_m$ , which does not use any knowledge on  $\sigma^2$ . This selection is done by

$$\hat{m} = \arg \min_{m \in \mathcal{M}_n} [\check{\gamma}_n(\hat{\sigma}_m^2) + \widetilde{\text{pen}}(m)]. \quad (5.31)$$

We denote by  $\hat{\sigma}_{\hat{m}}^2$  the resulting estimator and we need to determine the penalty  $\widetilde{\text{pen}}$  as for  $b$ . For simplicity, we use the same notation  $\hat{m}$  in (5.31) as in (5.20) although they are different. We can prove the following theorem.

**Theorem 5.9** *Assume that [A1]-[A7] hold. Consider collection [DP] with  $L_m = 1$  or [GP] with  $L_m$  given by (5.13) both with maximal dimension  $N_n \leq n\Delta/\ln^2(n)$ . Let*

$$\widetilde{\text{pen}}(m) \geq \tilde{\kappa} \sigma_1^4 \frac{(1 + L_m) D_m}{n},$$

where  $\tilde{\kappa}$  is a universal constant. Then, the estimator  $\hat{\sigma}_{\hat{m}}^2$  of  $\sigma^2$  with  $\hat{m}$  defined by (5.31) is such that

$$\mathbb{E}(\|\hat{\sigma}_{\hat{m}}^2 - \sigma_A^2\|_n^2) \leq C \inf_{m \in \mathcal{M}_n} (\|\sigma_m^2 - \sigma_A^2\|^2 + \widetilde{\text{pen}}(m)) + K' \Delta^2 + \frac{K''}{n}.$$

Analogous comments as those given for the drift can be made.

## 5.6.2 Proof of Proposition 5.8

Let us set

$$\check{\nu}_n(t) = \check{\nu}_n^{(1)}(t) + \check{\nu}_n^{(2)}(t) + \check{\nu}_n^{(3)}(t) \quad (5.32)$$

with

$$\check{\nu}_n^{(i)}(t) = \frac{1}{n} \sum_{k=1}^n t(X_{k\Delta}) V_{k\Delta}^{(i)}, \quad (5.33)$$

and

$$\check{\tau}_n(t) = \frac{1}{n} \sum_{k=1}^n t(X_{k\Delta}) \tau_{k\Delta}.$$

We begin with some lemmas. The first one concerns the remainder term.

**Lemma 5.10** *We have (see (5.27))*

$$\mathbb{E}\left(\frac{1}{n} \sum_{k=1}^n \tau_{k\Delta}^2\right) \leq K\Delta^2. \quad (5.34)$$

*Proof of Lemma 5.10.* We prove that  $\mathbb{E}[(\tau_{k\Delta}^{(i)})^2] \leq K_i\Delta^2$  for  $i = 1, 2, 3$ . Using [A5] and Lemma 5.1,

$$\begin{aligned} \mathbb{E}[(\tau_{k\Delta}^{(1)})^2] &\leq \mathbb{E}\left(\int_{k\Delta}^{(k+1)\Delta} b^2(X_s) ds\right)^2 \leq \Delta \mathbb{E}\left(\int_{k\Delta}^{(k+1)\Delta} b^4(X_s) ds\right) \\ &\leq \Delta^2 \mathbb{E}(b^4(X_0)) \leq c\Delta^2, \\ \mathbb{E}[(\tau_{k\Delta}^{(2)})^2] &\leq \frac{1}{\Delta^2} \left( \mathbb{E}\left(\int_{k\Delta}^{(k+1)\Delta} (b(X_s) - b(X_{k\Delta})) ds\right)^4 \right. \\ &\quad \left. \times \mathbb{E}\left(\int_{k\Delta}^{(k+1)\Delta} \sigma(X_s) dW_s\right)^4 \right)^{1/2} \end{aligned}$$

Using [A5], Lemma 5.1 and the Burkholder–Davis–Gundy inequality, we get

$$\mathbb{E}[(\tau_{k\Delta}^{(2)})^2] \leq c'\Delta^2.$$

Lastly, [A5] implies that  $|\psi(x)| \leq K(1 + x^2)$  (see (5.28)), hence

$$\begin{aligned} \mathbb{E}[(\tau_{k\Delta}^{(3)})^2] &\leq \frac{1}{\Delta} \mathbb{E}\left(\int_{k\Delta}^{(k+1)\Delta} ((k+1)\Delta - s)^2 \psi^2(X_s) ds\right) \\ &\leq \mathbb{E}(\psi^2(X_0)) \frac{\Delta^2}{3} \leq c''\Delta^2. \end{aligned}$$

Therefore (5.34) is proved.  $\square$

Now, we deal with the noise terms and show that  $i = 1$  gives the main term. In the statement below,  $K, K'$  denote constants which may vary from line to line.

**Lemma 5.11** 1. For  $S_m$  in collection [DP] or [GP],

$$\mathbb{E} \left( \sup_{t \in S_m, \|t\|_\pi = 1} (\check{\nu}_n^{(1)}(t))^2 \right) \leq K \frac{D_m}{n} \sigma_1^4.$$

2. Recall that  $S_n$  denotes the maximal space for both collections. For  $i = 2, 3$ ,

$$\mathbb{E} \left( \sup_{t \in S_n, \|t\|_\pi = 1} (\check{\nu}_n^{(i)}(t))^2 \right) \leq K \frac{\Delta N_n}{n} \leq K' \Delta^2. \quad (5.35)$$

*Proof of Lemma 5.11.* To study  $\check{\nu}_n^{(1)}(t)$ , we consider, as for the drift case, an orthonormal basis  $(\psi_\lambda, \lambda \in J_m)$  of  $S_m$  with respect to  $\mathbb{L}_\pi^2$ . So,

$$\mathbb{E} \left( \sup_{t \in S_m, \|t\|_\pi = 1} (\check{\nu}_n^{(1)}(t))^2 \right) \leq \sum_{\lambda \in J_m} \mathbb{E}((\check{\nu}_n^{(1)}(\psi_\lambda))^2).$$

Then, we use the fact that  $V_{k\Delta}^{(1)}$  is a martingale increment and obtain:

$$\mathbb{E}((\check{\nu}_n^{(1)}(\psi_\lambda))^2) = \frac{1}{n^2} \sum_{k=1}^n \mathbb{E}(\psi_\lambda^2(X_{k\Delta}) \mathbb{E}([V_{k\Delta}^{(1)}]^2 | \mathcal{F}_{k\Delta})).$$

Then,

$$\begin{aligned} \mathbb{E}((V_{k\Delta}^{(1)})^2 | \mathcal{F}_{k\Delta}) &\leq \frac{2}{\Delta^2} \left[ \mathbb{E} \left( \left( \int_{k\Delta}^{(k+1)\Delta} \sigma(X_s) dW_s \right)^4 | \mathcal{F}_{k\Delta} \right) \right. \\ &\quad \left. + \mathbb{E} \left( \left( \int_{k\Delta}^{(k+1)\Delta} \sigma^2(X_s) ds \right)^2 | \mathcal{F}_{k\Delta} \right) \right]. \end{aligned}$$

Using the Burkholder–Davis–Gundy inequality, we obtain:

$$\mathbb{E}((V_{k\Delta}^{(1)})^2 | \mathcal{F}_{k\Delta}) \leq C \sigma_1^4.$$

This gives the first part.

For the second part, note that the maximal space  $S_n$  is equipped with an orthonormal basis  $(\varphi_\lambda, \lambda \in \mathcal{L}_n)$  with respect to  $\mathbb{L}^2$  which satisfies for both collections (see (5.11) – (5.12))

$$\left\| \sum_{\lambda \in \mathcal{L}_n} \varphi_\lambda^2 \right\|_\infty \leq \Phi_0^2 N_n,$$

with  $\Phi_0^2 = r_{max} + 1$ . For  $i = 2, 3$ ,

$$\mathbb{E} \left( \sup_{t \in S_n, \|t\| \leq 1} (\check{\nu}_n^{(i)}(t))^2 \right) \leq \sum_{\lambda \in \mathcal{L}_n} \mathbb{E}((\check{\nu}_n^{(i)}(\varphi_\lambda))^2).$$

Since the martingale increments  $(V_{k\Delta}^{(i)})$  are uncorrelated, we have:

$$\mathbb{E}((\check{\nu}_n^{(i)}(\varphi_\lambda))^2) = \frac{1}{n^2} \sum_{k=1}^n \mathbb{E} \left( \varphi_\lambda^2(X_{k\Delta}) (V_{k\Delta}^{(i)})^2 \right).$$

Therefore, interchanging sums in  $\lambda$  and  $k$ , we get:

$$\sum_{\lambda \in \mathcal{L}_n} \mathbb{E}((\check{\nu}_n^{(i)}(\varphi_\lambda))^2) \leq \frac{\Phi_0^2 N_n}{n} \frac{1}{n} \sum_{k=1}^n \mathbb{E}((V_{k\Delta}^{(i)})^2). \quad (5.36)$$

Now,

$$\begin{aligned} \mathbb{E}((V_{k\Delta}^{(2)})^2) &= \frac{1}{\Delta^2} \mathbb{E}[(\sigma^2)'(X_0) \sigma(X_0)]^2 \int_{k\Delta}^{(k+1)\Delta} ((k+1)\Delta - s)^2 ds \\ &\leq C\Delta(1 + (\mathbb{E}(X_0))^4) \end{aligned}$$

and

$$\begin{aligned} \mathbb{E}((V_{k\Delta}^{(3)})^2) &= 4\mathbb{E}(b^2(X_{k\Delta}) \int_{k\Delta}^{(k+1)\Delta} \sigma^2(X_s) ds) \\ &\leq 4 \left( \mathbb{E}(b^4(X_{k\Delta})) \mathbb{E} \left[ \left( \int_{k\Delta}^{(k+1)\Delta} \sigma^2(X_s) ds \right)^2 \right] \right)^{1/2} \\ &\leq 4 (\mathbb{E}(b^4(X_0)) \mathbb{E}(\sigma^4(X_0)))^{1/2} \Delta \leq C\Delta(1 + \mathbb{E}(X_0^4)). \end{aligned}$$

Since  $\|t\|^2 \leq \|t\|_\pi^2 / \pi_0$ , we join the above bounds and (5.36) and obtain the first inequality in (5.35). Since  $N_n \leq n\Delta / \ln^2 n$ ,  $N_n\Delta/n \leq \Delta^2 / \ln^2 n$ . This gives the second inequality.  $\square$

Now, we can prove Proposition 5.8. As for the drift, the starting point is:

$$\check{\gamma}_n(t) - \check{\gamma}_n(\sigma^2) = \|\sigma^2 - t\|_n^2 + 2\check{\nu}_n(\sigma^2 - t) + 2\check{\tau}_n(\sigma^2 - t).$$

Introducing the orthogonal projection  $\sigma_m^2$  of  $\sigma^2$  on  $S_m$ , we have:

$$\check{\gamma}_n(\hat{\sigma}_m^2) - \check{\gamma}_n(\sigma^2) \leq \check{\gamma}_n(\sigma_m^2) - \check{\gamma}_n(\sigma^2).$$

After some computations analogous to those done for the drift study, we are led to the following inequality which holds on  $\Omega_n$  (see (5.24)):

$$\frac{1}{4} \|\hat{\sigma}_m^2 - \sigma_A^2\|_n^2 \leq \frac{7}{4} \|\sigma_m^2 - \sigma_A^2\|_n^2 + 8 \sup_{t \in B_m^\pi(0,1)} \check{\nu}_n^2(t) + \frac{8}{n} \sum_{k=1}^n \tau_{k\Delta}^2,$$

where  $B_m^\pi(0,1) = \{t \in S_m, \|t\|_\pi = 1\}$ . Now we apply Lemma 5.10 and Lemma 5.11. This yields the first three terms of the right-hand-side of (5.30).

The study on  $\Omega_n^c$  is the same as for  $b$  with the regression model  $U_{k\Delta} = \sigma^2(X_{k\Delta}) + \xi_{k\Delta}$ , where  $\xi_{k\Delta} = V_{k\Delta} + \tau_{k\Delta}$ . By standard inequalities,  $\mathbb{E}(\xi_{k\Delta}^4) \leq K\{\Delta^4 \mathbb{E}(b^8(X_0)) + \mathbb{E}(\sigma^8(X_0))\}$ . Hence,  $\mathbb{E}(\xi_{k\Delta}^4)$  is bounded. Moreover, using Lemma 5.4,  $\mathbb{P}(\Omega_n^c) \leq c/n^2$ .  $\square$

## 5.6.3 Proof of Theorem 5.9

This proof follows the same lines as the proof of Theorem 5.3. We start with a Bernstein-type inequality.

**Lemma 5.12** *Under the assumptions of Theorem 5.9,*

$$\mathbb{P}\left(\check{\nu}_n^{(1)}(t) \geq \epsilon, \|t\|_n^2 \leq v^2\right) \leq \exp\left(-Cn \frac{\epsilon^2/2}{2\sigma_1^4 v^2 + \epsilon \|t\|_\infty \sigma_1^2 v}\right)$$

and

$$\mathbb{P}\left(\check{\nu}_n^{(1)}(t) \geq v\sigma_1^2(2x)^{1/2} + \sigma_1^2 \|t\|_\infty x, \|t\|_n^2 \leq v^2\right) \leq \exp(-Cnx). \quad (5.37)$$

The proof that the first inequality implies the second one above is rather tricky and proved in Birgé and Massart (1998). Consequently, we just prove the first one.

*Proof of Lemma 5.12.* First we note that:

$$\begin{aligned} \mathbb{E}\left(e^{ut(X_{n\Delta})V_{n\Delta}^{(1)}} | \mathcal{F}_{n\Delta}\right) &= 1 + \sum_{p=2}^{+\infty} \frac{u^p}{p!} \mathbb{E}\left\{(t(X_{n\Delta})V_{n\Delta}^{(1)})^p | \mathcal{F}_{n\Delta}\right\} \\ &\leq 1 + \sum_{p=2}^{+\infty} \frac{u^p}{p!} |t(X_{n\Delta})|^p \mathbb{E}\left(|V_{n\Delta}^{(1)}|^p | \mathcal{F}_{n\Delta}\right). \end{aligned}$$

Next we apply successively the Minkowski inequality and the Burkholder–Davis–Gundy inequality with best constant (Proposition 4.2 of Barlow and Yor (1982)). For a continuous martingale  $(M_t)$ , with  $M_0 = 0$ , for  $k \geq 2$ ,  $M_t^* = \sup_{s \leq t} |M_s|$  satisfies  $\|M^*\|_k \leq ck^{1/2} \|\langle M \rangle^{1/2}\|_k$ , with  $c$  a universal constant. And we obtain:

$$\begin{aligned} \mathbb{E}(|V_{n\Delta}^{(1)}|^p | \mathcal{F}_{n\Delta}) &\leq \frac{2^{p-1}}{\Delta^p} \left\{ \mathbb{E}\left(\left|\int_{n\Delta}^{(n+1)\Delta} \sigma(X_s) dW_s\right|^{2p} | \mathcal{F}_{n\Delta}\right) \right. \\ &\quad \left. + \mathbb{E}\left(\left|\int_{n\Delta}^{(n+1)\Delta} \sigma^2(X_s) ds\right|^p | \mathcal{F}_{n\Delta}\right) \right\} \\ &\leq \frac{2^{p-1}}{\Delta^p} (c^{2p}(2p)^p \Delta^p \sigma_1^{2p} + \Delta^p \sigma_1^{2p}) \leq (2\sigma_1 c)^{2p} p^p. \end{aligned}$$

Therefore,

$$\mathbb{E}\left(e^{ut(X_{n\Delta})V_{n\Delta}^{(1)}} | \mathcal{F}_{n\Delta}\right) \leq 1 + \sum_{p=2}^{\infty} \frac{p^p}{p!} (4u\sigma_1^2 c^2)^p |t(X_{n\Delta})|^p.$$

Using  $p^p/p! \leq e^{p-1}$ , we find

$$\begin{aligned} \mathbb{E} \left( e^{ut(X_{n\Delta})V_{n\Delta}^{(1)}} | \mathcal{F}_{n\Delta} \right) &\leq 1 + e^{-1} \sum_{p=2}^{\infty} (4u\sigma_1^2 c^2 e)^p |t(X_{n\Delta})|^p \\ &\leq 1 + e^{-1} \frac{(4u\sigma_1^2 c^2 e)^2 t^2(X_{n\Delta})}{1 - (4u\sigma_1^2 c^2 e \|t\|_{\infty})}. \end{aligned}$$

Now, let us set

$$a = e(4\sigma_1^2 c^2)^2 \text{ and } b = 4\sigma_1^2 c^2 e \|t\|_{\infty}.$$

Since for  $x \geq 0$ ,  $1 + x \leq e^x$ , we get, for all  $u$  such that  $bu < 1$ ,

$$\mathbb{E} \left( e^{ut(X_{n\Delta})V_{n\Delta}^{(1)}} | \mathcal{F}_{n\Delta} \right) \leq 1 + \frac{au^2 t^2(X_{n\Delta})}{1 - bu} \leq \exp \left( \frac{au^2 t^2(X_{n\Delta})}{1 - bu} \right).$$

This can also be written:

$$\mathbb{E} \left( \exp \left( ut(X_{n\Delta})V_{n\Delta}^{(1)} - \frac{au^2 t^2(X_{n\Delta})}{1 - bu} \right) | \mathcal{F}_{n\Delta} \right) \leq 1.$$

Therefore, iterating conditional expectations yields

$$\mathbb{E} \left[ \exp \left\{ \sum_{k=1}^n \left( ut(X_{k\Delta})V_{k\Delta}^{(1)} - \frac{au^2 t^2(X_{k\Delta})}{1 - bu} \right) \right\} \right] \leq 1.$$

Then, we deduce that

$$\begin{aligned} &\mathbb{P} \left( \sum_{k=1}^n t(X_{k\Delta})V_{k\Delta}^{(1)} \geq n\epsilon, \|t\|_n^2 \leq v^2 \right) \\ &\leq e^{-nu\epsilon} \mathbb{E} \left\{ \mathbf{1}_{\|t\|_n^2 \leq v^2} \exp \left( u \sum_{k=1}^n t(X_{k\Delta})V_{k\Delta}^{(1)} \right) \right\} \\ &\leq e^{-nu\epsilon} \mathbb{E} \left[ \mathbf{1}_{\|t\|_n^2 \leq v^2} \exp \left\{ \sum_{k=1}^n \left( ut(X_{k\Delta})V_{k\Delta}^{(1)} - \frac{au^2 t^2(X_{k\Delta})}{1 - bu} \right) \right\} \right. \\ &\quad \left. \times e^{(au^2)/(1-bu) \sum_{k=1}^n t^2(X_{k\Delta})} \right] \\ &\leq e^{-nu\epsilon} e^{(nau^2 v^2)/(1-bu)} \mathbb{E} \left[ \exp \left\{ \sum_{k=1}^n \left( ut(X_{k\Delta})V_{k\Delta}^{(1)} - \frac{au^2 t^2(X_{k\Delta})}{1 - bu} \right) \right\} \right] \\ &\leq e^{-nu\epsilon} e^{(nau^2 v^2)/(1-bu)}. \end{aligned}$$

The inequality holds for any  $u$  such that  $bu < 1$ . In particular,  $u = \epsilon/(2av^2 + \epsilon b)$  gives  $-u\epsilon + au^2 v^2/(1 - bu) = -(1/2)(\epsilon^2/(2av^2 + \epsilon b))$  and therefore

$$\mathbb{P} \left( \sum_{k=1}^n t(X_{k\Delta})V_{k\Delta}^{(1)} \geq n\epsilon, \|t\|_n^2 \leq v^2 \right) \leq \exp \left( -n \frac{\epsilon^2/2}{2av^2 + \epsilon b} \right). \quad \square$$

We now finish the proof of Theorem 5.9. As for  $\hat{b}_{\hat{m}}$ , we introduce the additional penalty terms and obtain that the risk satisfies

$$\begin{aligned} & \mathbb{E}(\|\hat{\sigma}_{\hat{m}}^2 - \sigma_A^2\|_n^2 \mathbf{1}_{\Omega_n}) \\ & \leq 7\pi_1 \|\sigma_m^2 - \sigma_A^2\|^2 + 4\widehat{\text{pen}}(m) + 32\mathbb{E}\left(\sup_{t \in B_{m, \hat{m}}^\pi(0,1)} (\check{\nu}_n(t))^2 \mathbf{1}_{\Omega_n}\right) \\ & \quad - 4\mathbb{E}(\widehat{\text{pen}}(\hat{m})) + K'\Delta^2 \end{aligned} \quad (5.38)$$

where  $B_{m, m'}^\pi(0, 1) = \{t \in S_m + S_{m'}, \|t\|_\pi = 1\}$ . We use that

$$(\check{\nu}_n(t))^2 \leq 2[(\check{\nu}_n^{(1)}(t))^2 + (\check{\nu}_n^{(2)}(t) + \check{\nu}_n^{(3)}(t))^2].$$

By Lemma 5.11, since  $B_{m, m'}^\pi(0, 1) \subset \{t \in \mathcal{S}_n, \|t\|_\pi = 1\}$ ,

$$\mathbb{E}\left(\sup_{t \in B_{m, \hat{m}}^\pi(0,1)} (\check{\nu}_n^{(2)}(t) + \check{\nu}_n^{(3)}(t))^2\right) \leq K\Delta^2.$$

There remains the main term to study

$$\check{G}_m(m') = \sup_{t \in B_{m, m'}^\pi(0,1)} |\check{\nu}_n^{(1)}(t)|. \quad (5.39)$$

As for the drift, we write

$$\begin{aligned} \mathbb{E}(\check{G}_m^2(\hat{m})) & \leq \mathbb{E}[(\check{G}_m^2(\hat{m}) - \tilde{p}(m, \hat{m}))\mathbf{1}_{\Omega_n}]_+ + \mathbb{E}(\tilde{p}(m, \hat{m})) \\ & \leq \sum_{m' \in \mathcal{M}_n} \mathbb{E}[(\check{G}_m^2(m') - \tilde{p}(m, m'))\mathbf{1}_{\Omega_n}]_+ + \mathbb{E}(\tilde{p}(m, \hat{m})). \end{aligned}$$

Now we have the following statement.

**Proposition 5.13** *Under the assumptions of Theorem 5.9, for*

$$\tilde{p}(m, m') = \kappa\sigma_1^4 \frac{D_m + D_{m'}(1 + L_{m'})}{n} + K\Delta^2,$$

where  $\kappa$  is a numerical constant, we have

$$\mathbb{E}[(\check{G}_m^2(m') - \tilde{p}(m, m'))\mathbf{1}_{\Omega_n}]_+ \leq c\sigma_1^4 \frac{e^{-D_{m'}L_{m'}}}{n}.$$

The result of Proposition 5.13 is obtained from inequality (5.37) of Lemma 5.12 by a  $L_\pi^2 - L^\infty$  chaining technique. A description of this method, in a more general setting, is given in Propositions 2–4, pp. 282–287, in Comte (2001), Theorem 5 in Birgé and Massart (1998) and Proposition 7, Theorem 8 and Theorem 9 in Barron et al. (1999). For the sake of completeness and since the context is slightly different, we detail the proof in the Appendix, Section 5.9. Note that there is a difference between Propositions 5.6 and 5.13 which comes from the additional term  $\|t\|_\infty$  appearing in Lemma 5.12.



Choosing  $\widehat{\text{pen}}(m) \geq \tilde{\kappa}\sigma_1^4 D_m(1 + L_m)/n$  with  $\tilde{\kappa} = 16\kappa$ , we deduce from (5.38), Proposition 5.13 and  $D_m \leq N_n \leq n\Delta/\ln^2(n)$  that,

$$\begin{aligned} \mathbb{E}(\|\hat{\sigma}_m^2 - \sigma_A^2\|_n^2) &\leq 7\pi_1\|\sigma_m^2 - \sigma_A^2\|^2 + 8\widehat{\text{pen}}(m) + c\sigma_1^4 \sum_{m' \in \mathcal{M}_n} \frac{e^{-D_{m'}L_{m'}}}{n} \\ &\quad + K'\Delta^2 + \mathbb{E}(\|\hat{\sigma}_m^2 - \sigma_A^2\|_n^2 \mathbf{1}_{\Omega_n^c}). \end{aligned}$$

The bound for  $\mathbb{E}(\|\hat{\sigma}_m^2 - \sigma^2\|_n^2 \mathbf{1}_{\Omega_n^c})$  is the same as the one given in the end of the proof of Proposition 5.8. It is less than  $c/n$ . The result of Theorem 5.9 follows.  $\square$

## 5.7 Examples and practical implementation

In this section, we consider classical examples of diffusions for which an exact simulation of sample paths is possible and for which the estimation method has been implemented with [GP]. For exact simulation of sample paths, when it is not directly possible, we have in view the retrospective exact simulation algorithms proposed by Beskos, Papaspiliopoulos, and Roberts (2006) and Beskos and Roberts (2005). Models of Families 1 and 2 below can be simulated by the algorithm EA1. Among the assumptions, requiring that  $\sigma$  be bounded is rather stringent and not always satisfied in our examples. The other assumptions hold. More details may be found in Comte et al. (2006, 2007)

### 5.7.1 Examples of diffusions

#### Family 1

First, we consider (5.1) with

$$b(x) = -\theta x, \quad \sigma(x) = c(1 + x^2)^{1/2}.$$

Standard computations of the scale and speed densities show that the model is positive recurrent for  $\theta + c^2/2 > 0$ . In this case, its stationary distribution has density

$$\pi(x) \propto \frac{1}{(1 + x^2)^{1+\theta/c^2}}.$$

If  $X_0 = \eta$  has distribution  $\pi(x)dx$ , then, setting  $\nu = 1 + 2\theta/c^2$ ,  $\nu^{1/2}\eta$  has Student distribution  $t(\nu)$ . This distribution satisfies the moment condition [A6] for  $2\theta/c^2 > 7$ . See Figure 5.1 for the estimation of  $b$  and  $\sigma^2$  in this case.

Then, we consider  $F_1(x) = \int_0^x 1/(c(1 + x^2)^{1/2})dx = \arg \sinh(x)/c$ . By the Itô formula,  $\xi_t = F_1(X_t)$  is solution of a stochastic differential equation with  $\sigma(\xi) = 1$  and

$$b(\xi) = -(\theta/c + c/2) \tanh(c\xi).$$

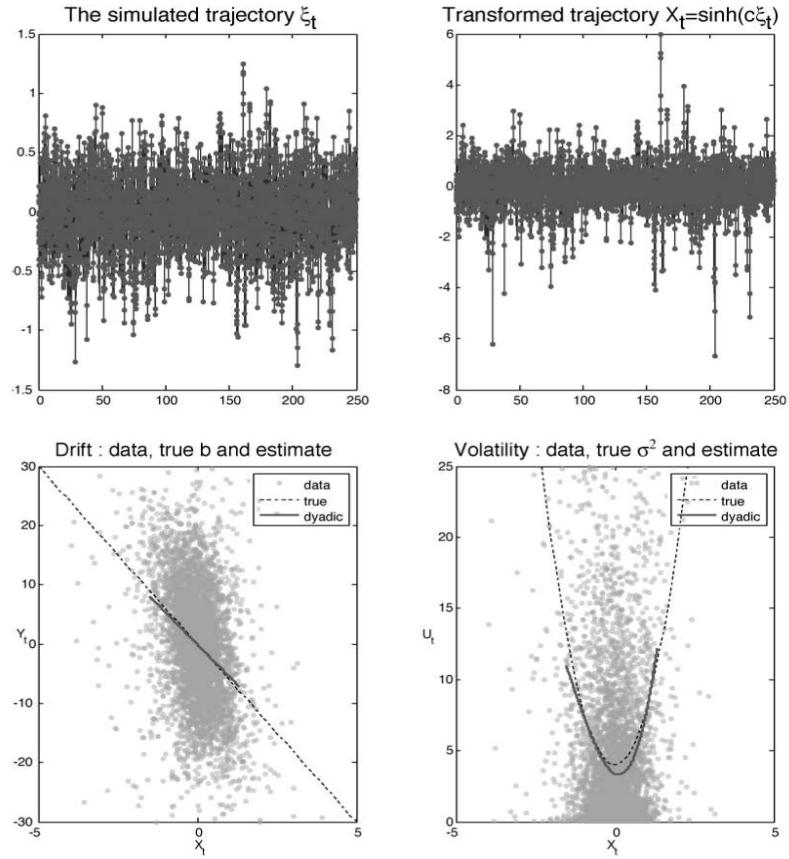


Figure 5.1 First example:  $dX_t = -\theta X_t dt + c\sqrt{1 + X_t^2} dW_t$ ,  $n = 5000$ ,  $\Delta = 1/20$ ,  $\theta = 2$ ,  $c = 1$ , dotted line: true function, full line: estimated function.

Assumptions [A1] – [A3] and [A5] hold for  $(\xi_t)$  with  $\xi_0 = F_1(X_0)$ . Moreover,  $(\xi_t)$  satisfies the conditions of Proposition 1 in Pardoux and Veretennikov (2001) implying that  $(\xi_t)$  is exponentially  $\beta$ -mixing and has moments of any order. Hence, [A4] and [A6] hold. See Figure 5.2 for the estimation of  $b$  and  $\sigma^2$  in this case.

Since  $X_t = F_1^{-1}(\xi_t)$ , this process is also  $\beta$ -mixing. It satisfies all assumptions except that  $\sigma^2(x)$  is not bounded from above.

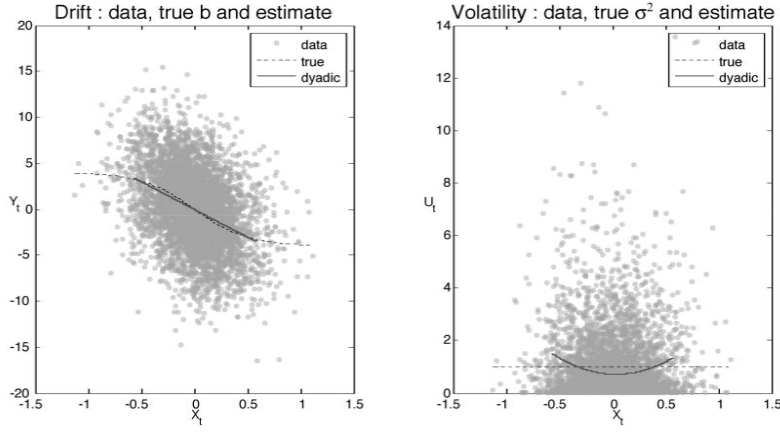


Figure 5.2 *Second example:  $d\xi_t = -(\theta/c + c/2) \tanh(c\xi_t) + dW_t$ ,  $n = 5000$ ,  $\Delta = 1/20$ ,  $\theta = 6$ ,  $c = 2$ , dotted line: true function, full line: estimated function.*

### Family 2

For the second family of models, we consider a process  $(\xi_t)$  with diffusion coefficient  $\sigma(\xi) = 1$  and drift

$$b(\xi) = -\theta \frac{\xi}{(1 + c^2 \xi^2)^{1/2}}, \quad (5.40)$$

(see Barndorff-Nielsen (1978)). The model is positive recurrent on  $\mathbb{R}$  for  $\theta > 0$ . Its stationary distribution is a hyperbolic distribution given by

$$\pi(\xi) d\xi \propto \exp\left(-2 \frac{\theta}{c^2} (1 + c^2 \xi^2)^{1/2}\right).$$

Assumptions [A1] – [A3], [A5] – [A6] hold for this model. For [A4], we apply Proposition 1 of Pardoux and Veretennikov (2001).

Next, we consider  $X_t = F_2(\xi_t) = \arg \sinh(c\xi_t)$  which satisfies a stochastic differential equation with coefficients:

$$b(x) = -\left(\theta + \frac{c^2}{2 \cosh(x)}\right) \frac{\sinh(x)}{\cosh^2(x)}, \quad \sigma(x) = \frac{c}{\cosh(x)}.$$

The process  $(X_t)$  is exponentially  $\beta$ -mixing as  $(\xi_t)$ . The diffusion coefficient  $\sigma(x)$  has an upper bound. See Figure 5.3 for the estimation of  $b$  and  $\sigma^2$  in this case.

To obtain a different shape for the diffusion coefficient, showing two bumps,

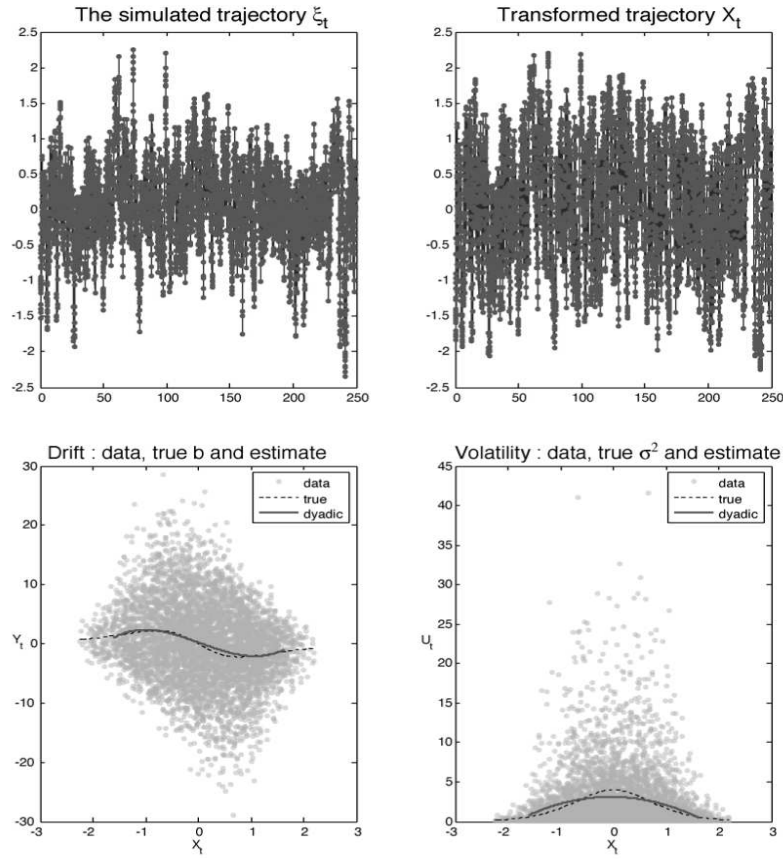


Figure 5.3 *Third example,  $dX_t = -[\theta + c^2/(2 \cosh(X_t))] (\sinh(X_t)/\cosh^2(X_t))dt + (c/\cosh(X_t))dW_t$ ,  $n = 5000$ ,  $\Delta = 1/20$ ,  $\theta = 3$ ,  $c = 2$ , dotted line: true function, full line: estimated function.*

we consider  $X_t = G(\xi_t) = \arg \sinh(\xi_t - 5) + \arg \sinh(\xi_t + 5)$ . The function  $G(\cdot)$  is invertible and its inverse has the following explicit expression,

$$G^{-1}(x) = \frac{1}{2^{1/2} \sinh(x)} [49 \sinh^2(x) + 100 + \cosh(x)(\sinh^2(x) - 100)]^{1/2}.$$

The diffusion coefficient of  $(X_t)$  is given by

$$\sigma(x) = \frac{1}{(1 + (G^{-1}(x) - 5)^2)^{1/2}} + \frac{1}{(1 + (G^{-1}(x) + 5)^2)^{1/2}}.$$

The drift is given by  $G'(G^{-1}(x))b(G^{-1}(x)) + \frac{1}{2}G''(G^{-1}(x))$  with  $b$  given in (5.40). See Figure 5.4 for the estimation of  $b$  and  $\sigma^2$  in this case.

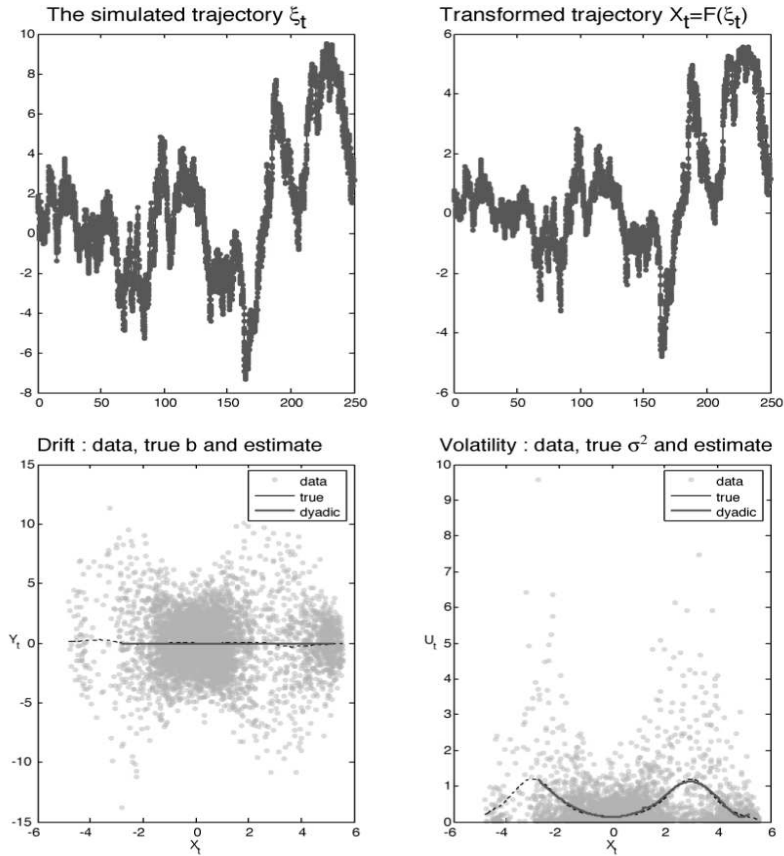


Figure 5.4 *Fourth example, the two-bumps diffusion coefficient  $X_t = G(\xi_t)$ ,  $d\xi_t = -\theta\xi_t/\sqrt{1+c^2\xi_t^2}dt + dW_t$ ,  $G(x) = \arg \sinh(x-5) + \arg \sinh(x+5)$ ,  $n = 5000$ ,  $\Delta = 1/20$ ,  $\theta = 1$ ,  $c = 10$ , dotted line: true function, full line: estimated function.*

### Family 3

Consider  $Y_t$  a stationary Ornstein-Uhlenbeck process given by  $dY_t = -\theta Y_t dt + c dW_t$  with  $\theta > 0$  and  $Y_0 \rightsquigarrow \mathcal{N}(0, c^2/(2\theta))$ . The  $\beta$ -mixing coefficient of  $(Y_t)$

can be evaluated using the exact formula (5.2). This gives a direct proof of (5.3).

**Proposition 5.14** *The  $\beta$ -mixing coefficient of  $(Y_t)$  satisfies*

$$\beta_Y(t) \leq \frac{\exp(-\theta t)}{2(1 - \exp(-\theta t))}.$$

*Proof.* We use the expansion of the transition density  $p_t(x, y)$  of  $(Y_t)$  in terms of the sequence of eigenvalues and eigenfunctions of the infinitesimal generator  $Lf(y) = \frac{\sigma^2}{2}f''(y) - \theta yf'(y)$ . For this, we refer e.g. to Karlin and Taylor (1981, p.333). To simplify notations, we assume that  $\sigma^2/2\theta = 1$  so that the stationary distribution of  $(Y_t)$  is  $\pi(y)dy = \mathcal{N}(0, 1)$ . Let us now consider the  $n$ -th Hermite polynomial given, for  $n = 0, 1, \dots$ , by:

$$H_n(x) = \frac{(-1)^n}{\sqrt{n!}} \exp(x^2/2) \frac{d^n}{dx^n} [\exp(-x^2/2)].$$

As defined above, this sequence is an orthonormal basis of  $L^2(\pi)$  and satisfies, for all  $n \geq 0$ ,  $LH_n = -n\theta H_n$ , i.e.,  $H_n$  is the eigenfunction associated with the eigenvalue  $-n\theta$  of  $L$ . This gives the following expansion:

$$p_t(x, y) = \pi(y) \sum_{n=0}^{+\infty} \exp(-n\theta t) H_n(x) H_n(y).$$

Since  $H_0(x) = 1$  and the  $H_n$  have  $L^2(\pi)$ -norm equal to 1, we get

$$\begin{aligned} & \|p_t(x, y)dy - \pi(y)dy\|_{TV} \\ &= \frac{1}{2} \int_{\mathbb{R}} |p_t(x, y) - \pi(y)| dy \\ &\leq \frac{1}{2} \sum_{n=1}^{+\infty} \exp(-n\theta t) |H_n(x)| \int_{\mathbb{R}} |H_n(y)| \pi(y) dy \\ &\leq \frac{1}{2} \sum_{n=1}^{+\infty} \exp(-n\theta t) |H_n(x)|. \end{aligned}$$

Integrating w.r.t.  $\pi(x)dx$  and repeating the same tool, we obtain:

$$\beta_Y(t) \leq \frac{1}{2} \sum_{n=1}^{+\infty} \exp(-n\theta t) = \frac{\exp(-\theta t)}{2(1 - \exp(-\theta t))}.$$

□

The interest of this proof is that it can be mimicked for all models for which the infinitesimal generator has a discrete spectrum with explicit eigenfunctions and eigenvalues.

Now, we consider  $X_t = \tanh(Y_t)$ . By the Itô formula, we get that  $(X_t)$  has coefficients

$$b(x) = -(1-x^2) \left[ c^2 x + \frac{\theta}{2} \ln \left( \frac{1+x}{1-x} \right) \right], \quad \sigma(x) = c(1-x^2).$$

Assumptions [A1] – [A6] are satisfied for  $(X_t)$ .

Finally, we consider

$$dX_t = \left[ \frac{dc^2}{4} - \theta X_t \right] dt + c\sqrt{X_t} dW_t.$$

With  $d \geq 2$  integer,  $(X_t)$  has the distribution of  $\sum_{i=1}^d Y_{i,t}^2$  where  $(Y_{i,t})$  are i.i.d. Ornstein-Uhlenbeck processes as above. The process  $(X_t)$  satisfies all assumptions except that its diffusion coefficient is not bounded.

### 5.7.2 Calibrating the penalties

It is not easy to calibrate the penalties. The method is studied in full details in Comte and Rozenholc (2004). Implementation with [DP] is done on the above examples in Comte et al. (2007) and for [GP] in Comte et al. (2006). We only give here a brief description.

For collection [GP], the drift penalty ( $i = 1$ ) and the diffusion penalty ( $i = 2$ ) are given by

$$2 \frac{\hat{s}_i^2}{n} \left( d - 1 + \ln \left( \frac{d_{max} - 1}{d - 1} \right) + \ln^{2.5}(d) + \sum_{j=1}^d (r_j + \ln^{2.5}(r_j + 1)) \right).$$

Moreover,  $d_{max} = \lceil n\Delta / \ln^{1.5}(n) \rceil$ ,  $r_{max} = 5$ . The constants  $\kappa$  and  $\tilde{\kappa}$  in both drift and diffusion penalties have been set equal to 2. The term  $\hat{s}_1^2$  replaces  $\sigma_1^2/\Delta$  for the estimation of  $b$  and  $\hat{s}_2^2$  replaces  $\sigma_1^4$  for the estimation of  $\sigma^2$ . Let us first explain how  $\hat{s}_2^2$  is obtained. We run once the estimation algorithm of  $\sigma^2$  with a preliminary penalty where  $\hat{s}_2^2$  is taken equal to  $2 \max_m (\check{\gamma}_n(\hat{\sigma}_m^2))$ . This gives a preliminary estimator  $\tilde{\sigma}_0^2$ . Now, we take  $\hat{s}_2$  equal to twice the 99.5%-quantile of  $\tilde{\sigma}_0^2$ . The use of the quantile here is to avoid extreme values. We get  $\tilde{\sigma}^2$ . We use this estimate and set  $\hat{s}_1^2 = \max_{1 \leq k \leq n} (\tilde{\sigma}^2(X_{k\Delta}))/\Delta$  for the penalty of  $b$ . In all the examples, parameters have been chosen in the admissible range of ergodicity. The sample size  $n = 5000$  and the step  $\Delta = 1/20$  are in accordance with the asymptotic context (great  $n$ 's and small  $\Delta$ 's).

### 5.8 Bibliographical remarks

Non-parametric estimation of the coefficients of diffusion processes has been widely investigated in the last decades. There are two first reference papers which are only devoted to drift estimation. One is by Banon (1978), who uses a spectral approach and a continuous time observation of the sample path. The other one is by Tuan (1981) who constructs and studies kernel estimators of the drift based on a continuous time observation of the sample path and also on a discrete observation of the sample path for an ergodic diffusion process.

More recently, several authors have considered drift estimation based on a continuous time observation of the sample path for ergodic models. Asymptotic results are given as the length of the observation time interval tends to infinity (Prakasa Rao (2010), Spokoiny (2000), Kutoyants (2004) or Dalalyan (2005)).

Discrete sampling of observations has also been investigated, with different asymptotic frameworks, implying different statistical strategies. It is now classical to distinguish between low-frequency and high-frequency data. In the former case, observations are taken at regularly spaced instants with fixed sampling interval  $\Delta$  and the asymptotic framework is that the number of observations tends to infinity. Then, only ergodic models are usually considered. Parametric estimation in this context has been studied by Bibby and Sørensen (1995), Kessler and Sørensen (1999), see also Bibby, Jacobsen, and Sørensen (2009). A non-parametric approach using spectral methods is investigated in Gobet, Hoffmann, and Reiß (2004), where non-standard non-parametric rates are exhibited.

In high-frequency data, the sampling interval  $\Delta = \Delta_n$  between two successive observations is assumed to tend to zero as the number of observations  $n$  tends to infinity. Taking  $\Delta_n = 1/n$ , so that the length of the observation time interval  $n\Delta_n = 1$  is fixed, can only lead to estimating the diffusion coefficient consistently with no need of ergodicity assumptions. This is done by Hoffmann (1999) who generalizes results by Jacod (2000), Florens-Zmirou (1993) and Genon-Catalot, Larédo, and Picard (1992).

Now, estimating both drift and diffusion coefficients requires that the sampling interval  $\Delta_n$  tends to zero while  $n\Delta_n$  tends to infinity. For ergodic diffusion models, Hoffmann (1999) proposes non-parametric estimators using projections on wavelet bases together with adaptive procedures. He exhibits minimax rates and shows that his estimators automatically reach these optimal rates up to logarithmic factors. Hence, Hoffmann's paper gives the benchmark for studying non-parametric estimation in this framework and assumptions. Nevertheless, Hoffmann's estimators are based on computations of some random times which make them difficult to implement.

Finally, let us mention that Bandi and Phillips (2003) also consider the same



asymptotic framework but with nonstationary diffusion processes: they study kernel estimators using local time estimations and random normalization.

### 5.9 Appendix. Proof of Proposition 5.13

The proof relies on the following Lemma (Lemma 9 in Barron et al. (1999)):

**Lemma 5.15** *Let  $\mu$  be a positive measure on  $[0, 1]$ . Let  $(\psi_\lambda)_{\lambda \in \Lambda}$  be a finite orthonormal system in  $\mathbb{L}_2 \cap \mathbb{L}_\infty(\mu)$  with  $|\Lambda| = D$  and  $\bar{S}$  be the linear span of  $\{\psi_\lambda\}$ . Let*

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

*For any positive  $\delta$ , one can find a countable set  $T \subset \bar{S}$  and a mapping  $p$  from  $\bar{S}$  to  $T$  with the following properties:*

- *for any ball  $\mathcal{B}$  with radius  $\sigma \geq 5\delta$ ,*

$$|T \cap \mathcal{B}| \leq (B'\sigma/\delta)^D \text{ with } B' < 5.$$

- *$\|u - p(u)\|_\mu \leq \delta$  for all  $u$  in  $\bar{S}$ , and*

$$\sup_{u \in p^{-1}(t)} \|u - t\|_\infty \leq \bar{r}\delta, \text{ for all } t \text{ in } T.$$

To use this lemma, the main difficulty is often to evaluate  $\bar{r}$  in the different contexts. In our problem, the measure  $\mu$  is  $\pi$ . We consider a collection of models  $(S_m)_{m \in \mathcal{M}_n}$  which can be [DP] or [GP]. Recall that  $B_{m,m'}^\pi(0, 1) = \{t \in S_m + S_{m'}, \|t\|_\pi = 1\}$ . We have to compute  $\bar{r} = \bar{r}_{m,m'}$  corresponding to  $\bar{S} = S_m + S_{m'}$ . We denote by  $D(m, m') = \dim(S_m + S_{m'})$ .

Collection [DP]–  $S_m + S_{m'} = S_{\max(m, m')}$ ,  $D(m, m') = \max(D_m, D_{m'})$ , an orthonormal  $\mathbb{L}_2(\pi)$ -basis  $(\psi_\lambda)_{\lambda \in \Lambda(m, m')}$  can be built by orthonormalisation, on each sub-interval, of  $(\varphi_\lambda)_{\lambda \in \Lambda(m, m')}$ . Then

$$\begin{aligned} & \sup_{\beta \neq 0} \frac{\|\sum_{\lambda \in \Lambda(m, m')} \beta_\lambda \psi_\lambda\|_\infty}{|\beta|_\infty} \\ & \leq \left\| \sum_{\lambda \in \Lambda(m, m')} \psi_\lambda \right\|_\infty \leq (r_{\max} + 1) \sup_{\lambda \in \Lambda(m, m')} \|\psi_\lambda\|_\infty \\ & \leq (r_{\max} + 1)^{3/2} \sqrt{D(m, m')} \sup_{\lambda \in \Lambda(m, m')} \|\psi_\lambda\| \\ & \leq (r_{\max} + 1)^{3/2} \sqrt{D(m, m')} \sup_{\lambda \in \Lambda(m, m')} \|\psi_\lambda\|_\pi / \sqrt{\pi_0} \\ & \leq (r_{\max} + 1)^{3/2} \sqrt{D(m, m')/\pi_0}. \end{aligned}$$

Thus here  $\bar{r}_{m,m'} \leq (r_{max} + 1)^{3/2} / \sqrt{\pi_0}$ .

Collection [GP]– Here we have  $\bar{r}_{m,m'} \leq [(r_{max} + 1)\sqrt{N_n}] / \sqrt{D(m, m')\pi_0}$ .

We now prove Proposition 5.13. We apply Lemma 5.15 to the linear space  $S_m + S_{m'}$  of dimension  $D(m, m')$  and norm connection measured by  $\bar{r}_{m,m'}$  bounded above. We consider  $\delta_k$ -nets,  $T_k = T_{\delta_k} \cap B_{m,m'}^\pi(0, 1)$ , with  $\delta_k = \delta_0 2^{-k}$  with  $\delta_0 \leq 1/5$ , to be chosen later and we set

$$H_k = \ln(|T_k|) \leq D(m, m') \ln(5/\delta_k) = D(m, m') [k \ln(2) + \ln(5/\delta_0)]. \quad (5.41)$$

Given some point  $u \in B_{m,m'}^\pi(0, 1)$ , we can find a sequence  $\{u_k\}_{k \geq 0}$  with  $u_k \in T_k$  such that  $\|u - u_k\|_\pi^2 \leq \delta_k^2$  and  $\|u - u_k\|_\infty \leq \bar{r}_{m,m'} \delta_k$ . Thus we have the following decomposition that holds for any  $u \in B_{m,m'}^\pi(0, 1)$ ,

$$u = u_0 + \sum_{k=1}^{\infty} (u_k - u_{k-1}).$$

Clearly  $\|u_0\|_\pi \leq 1$ ,  $\|u_0\|_\infty \leq \bar{r}_{(m,m')}$  and for all  $k \geq 1$ ,  $\|u_k - u_{k-1}\|_\pi^2 \leq 2(\delta_k^2 + \delta_{k-1}^2) = 5\delta_{k-1}^2/2$  and  $\|u_k - u_{k-1}\|_\infty \leq 3\bar{r}_{(m,m')} \delta_{k-1}/2$ . In the sequel we denote by  $\mathbb{P}_n(\cdot)$  the measure  $\mathbb{P}(\cdot \cap \Omega_n)$ , see (5.24), (actually only the inequality  $\|t\|_n^2 \leq \frac{3}{2}\|t\|_\pi^2$  holding for any  $t \in S_m + S_{m'}$  is required).

Let  $(\eta_k)_{k \geq 0}$  be a sequence of positive numbers that will be chosen later on and  $\eta$  such that  $\eta_0 + \sum_{k \geq 1} \eta_k \leq \eta$ . Recall that  $\check{\nu}_n^{(1)}$  is defined by (5.27) – (5.32) – (5.33). We have

$$\begin{aligned} & \mathbb{P}_n \left[ \sup_{u \in B_{m,m'}^\pi(0,1)} \check{\nu}_n^{(1)}(u) > \eta \right] \\ &= \mathbb{P}_n \left[ \exists (u_k)_{k \in \mathbb{N}} \in \prod_{k \in \mathbb{N}} T_k / \right. \\ & \quad \left. \check{\nu}_n^{(1)}(u_0) + \sum_{k=1}^{+\infty} \check{\nu}_n^{(1)}(u_k - u_{k-1}) > \eta_0 + \sum_{k \geq 1} \eta_k \right] \\ &\leq \mathbb{P}_1 + \mathbb{P}_2 \end{aligned}$$

where

$$\begin{aligned} \mathbb{P}_1 &= \sum_{u_0 \in T_0} \mathbb{P}_n(\check{\nu}_n^{(1)}(u_0) > \eta_0), \\ \mathbb{P}_2 &= \sum_{k=1}^{\infty} \sum_{\substack{u_{k-1} \in T_{k-1} \\ u_k \in T_k}} \mathbb{P}_n(\check{\nu}_n^{(1)}(u_k - u_{k-1}) > \eta_k). \end{aligned}$$

Then using inequality (5.37) of Lemma 5.12 and (5.41), we straightforwardly infer that  $\mathbb{P}_1 \leq \exp(H_0 - Cnx_0)$  and  $\mathbb{P}_2 \leq \sum_{k \geq 1} \exp(H_{k-1} + H_k - Cnx_k)$

if we choose

$$\begin{cases} \eta_0 = \sigma_1^2(\sqrt{3x_0} + \bar{r}_{(m,m')}x_0) \\ \eta_k = (\sigma_1^2/\sqrt{2})\delta_{k-1}(\sqrt{15x_k} + 3\bar{r}_{(m,m')}x_k). \end{cases}$$

Fix  $\tau > 0$  and choose  $x_0$  such that

$$Cnx_0 = H_0 + L_{m'}D_{m'} + \tau$$

and for  $k \geq 1$ ,  $x_k$  such that

$$Cnx_k = H_{k-1} + H_k + kD_{m'} + L_{m'}D_{m'} + \tau.$$

If  $D_{m'} \geq 1$ , we infer that

$$\begin{aligned} \mathbf{P}_n \left( \sup_{t \in B_{m,m'}^\pi(0,1)} \check{\nu}_n^{(1)}(t) > \eta_0 + \sum_{k \geq 1} \eta_k \right) \\ \leq e^{-L_{m'}D_{m'} - \tau} \left( 1 + \sum_{k=1}^{\infty} e^{-kD_{m'}} \right) \leq 1.6e^{-L_{m'}D_{m'} - \tau}. \end{aligned}$$

Now, it remains to compute  $\sum_{k \geq 0} \eta_k$ . We note that  $\sum_{k=0}^{\infty} \delta_k = \sum_{k=0}^{\infty} k\delta_k = 2\delta_0$ . This implies

$$\begin{aligned} x_0 + \sum_{k=1}^{\infty} \delta_{k-1}x_k \\ \leq \left[ \ln(5/\delta_0) + \delta_0 \sum_{k=1}^{\infty} 2^{-(k-1)} [(2k-1)\ln(2) + 2\ln(5/\delta_0) + k] \right] \\ \times \frac{D(m,m')}{nC} \\ + \left( 1 + \delta_0 \sum_{k \geq 1} 2^{-(k-1)} \right) \left( \frac{L_{m'}D_{m'}}{nC} + \frac{\tau}{nC} \right) \\ \leq a(\delta_0) \frac{D(m,m')}{n} + \left( \frac{1+2\delta_0}{C} \right) \left( \frac{L_{m'}D_{m'}}{n} + \frac{\tau}{n} \right), \end{aligned} \quad (5.42)$$

where  $Ca(\delta_0) = \ln(5/\delta_0) + \delta_0(4 \ln(5/\delta_0) + 6 \ln(2) + 4)$ . This leads to

$$\begin{aligned}
& \left( \sum_{k=0}^{\infty} \eta_k \right)^2 \\
& \leq \frac{\sigma_1^4}{2} \left[ \sqrt{2} (\sqrt{3x_0} + \bar{r}_{m,m'} x_0) + \sum_{k=1}^{\infty} \delta_{k-1} (\sqrt{15x_k} + 3\bar{r}_{m,m'} x_k) \right]^2 \\
& \leq \frac{\sigma_1^4}{2} \left[ \left( \sqrt{6x_0} + \sum_{k=1}^{\infty} \delta_{k-1} \sqrt{15x_k} \right) + \bar{r}_{m,m'} \left( \sqrt{2x_0} + 3 \sum_{k=1}^{\infty} \delta_{k-1} x_k \right) \right]^2 \\
& \leq 15\sigma_1^4 \left[ \left( \sqrt{x_0} + \sum_{k=1}^{\infty} \delta_{k-1} \sqrt{x_k} \right)^2 + \bar{r}_{m,m'}^2 \left( x_0 + \sum_{k=1}^{\infty} \delta_{k-1} x_k \right)^2 \right] \\
& \leq 15\sigma_1^4 \left[ \left( 1 + \sum_{k=1}^{\infty} \delta_{k-1} \right) \left( x_0 + \sum_{k=1}^{\infty} \delta_{k-1} x_k \right) \right. \\
& \quad \left. + \bar{r}_{m,m'}^2 \left( x_0 + \sum_{k=1}^{\infty} \delta_{k-1} x_k \right)^2 \right].
\end{aligned}$$

Now, fix  $\delta_0 \leq 1/5$  (say,  $\delta_0 = 1/10$ ) and use the bound (5.42). The bound for  $(\sum_{k=0}^{+\infty} \eta_k)^2$  is less than a quantity proportional to

$$\begin{aligned}
& \sigma_1^4 \left[ \frac{D(m, m')}{n} + \frac{L_{m'} D_{m'}}{n} + \bar{r}_{m,m'}^2 \left( \frac{D(m, m')}{n} + \frac{L_{m'} D_{m'}}{n} \right)^2 \right. \\
& \quad \left. + \frac{\tau}{n} + \bar{r}_{m,m'}^2 \frac{\tau^2}{n^2} \right].
\end{aligned}$$

Now in the case of collection [DP], we have  $L_m = 1$ ,  $\bar{r}_{m,m'}$  is bounded uniformly with respect to  $m$  and  $m'$  and  $(D(m, m')/n)^2 \leq (N_n/n)^2 \leq \Delta^2/\ln^4(n)$  with  $N_n \leq n\Delta/\ln^2(n)$ . Thus the bound for  $(\sum \eta_k)^2$  reduces to

$$C' \sigma_1^4 \left[ \frac{D(m, m')}{n} + (1 + r_{max})^3 \Delta^2/\pi_0 + \frac{\tau}{n} + \bar{r}_{m,m'}^2 \frac{\tau^2}{n^2} \right].$$

Next, for collection [GP], we use that  $L_m \leq c \ln(n)$ ,  $\bar{r}_{m,m'}^2 \leq (r_{max} + 1)^3 N_n / (\pi_0 D(m, m'))$  and  $N_n \leq n \Delta / \ln^2(n)$  to obtain the bound

$$\begin{aligned}
& \bar{r}_{m,m'}^2 \left( \frac{D(m, m')}{n} + \frac{L_{m'} D_{m'}}{n} \right)^2 \\
& \leq (r_{max} + 1)^3 \frac{N_n}{\pi_0 D(m, m')} \frac{D(m, m')^2}{n^2} (1 + \ln(n))^2 \\
& \leq (r_{max} + 1)^3 \frac{N_n D(m, m')}{\pi_0 n^2} (1 + \ln(n))^2 \\
& \leq (r_{max} + 1)^3 \frac{N_n^2}{\pi_0 n^2} (1 + \ln(n))^2 \leq 2(r_{max} + 1)^3 \Delta^2 / \pi_0.
\end{aligned}$$

Thus, the bound for  $(\sum \eta_k)^2$  is proportional to

$$\sigma_1^4 \left[ \frac{D(m, m')}{n} + \frac{L_{m'} D_{m'}}{n} + 2(r_{max} + 1)^3 \Delta^2 / \pi_0 + \frac{\tau}{n} + \bar{r}_{m,m'}^2 \frac{\tau^2}{n^2} \right].$$

This term defines  $\tilde{p}(m, m')$  as given in Proposition 5.13.

We obtain, for  $K = (r_{max} + 1)^3 / \pi_0$ ,

$$\begin{aligned}
& \mathbb{P}_n \left[ \sup_{u \in B_{m,m'}^\pi(0,1)} [\check{\nu}_n^{(1)}(u)]^2 > \kappa \sigma_1^4 \left( \frac{D_m + D_{m'}(1 + L_{m'})}{n} \right. \right. \\
& \quad \left. \left. + K \Delta^2 + 2 \left( \frac{\tau}{n} \vee 2 \bar{r}_{m,m'}^2 \frac{\tau^2}{n^2} \right) \right) \right] \\
& \leq \mathbb{P}_n \left[ \sup_{u \in B_{m,m'}^\pi(0,1)} [\check{\nu}_n^{(1)}(u)]^2 > \eta^2 \right] \\
& \leq 2 \mathbb{P}_n \left[ \sup_{u \in B_{m,m'}^\pi(0,1)} \check{\nu}_n^{(1)}(u) > \eta \right] \\
& \leq 3.2 e^{-L_{m'} D_{m'} - \tau}
\end{aligned}$$

so that, reminding that  $\check{G}_m(m')$  is defined by (5.39),

$$\begin{aligned}
& \mathbb{E} \left[ \left( \check{G}_m^2(m') - \kappa \sigma_1^4 \frac{D_m + D_{m'}(1 + L_{m'})}{n} + K \Delta^2 \right)_+ \mathbf{1}_{\Omega_n} \right] \\
& \leq \int_0^\infty \mathbb{P}_n \left( \check{G}_m^2(m') > \kappa \sigma_1^4 \frac{D_m + D_{m'}(1 + L_{m'})}{n} + K \Delta^2 + \tau \right) d\tau \\
& \leq e^{-L_{m'} D_{m'}} \left( \int_{2\kappa \sigma_1^4 / \bar{r}_{(m, m')}^2}^\infty e^{-n\tau / (2\kappa \sigma_1^4)} d\tau \right. \\
& \quad \left. + \int_0^{2\kappa \sigma_1^4 / \bar{r}_{m, m'}^2} e^{-n\sqrt{\tau} / (2\sqrt{\kappa} \bar{r}_{m, m'} \sigma_1^2)} d\tau \right) \\
& \leq e^{-L_{m'} D_{m'}} \frac{2\kappa \sigma_1^4}{n} \left( \int_0^\infty e^{-v} dv + \frac{2\bar{r}_{m, m'}^2}{n} \int_0^\infty e^{-\sqrt{v}} dv \right) \\
& \leq e^{-L_{m'} D_{m'}} \frac{2\kappa \sigma_1^4}{n} \left( 1 + \frac{4\bar{r}_{m, m'}^2}{n} \right) \leq \kappa' e^{-L_{m'} D_{m'}} \frac{\sigma_1^4}{n}
\end{aligned}$$

which ends the proof.  $\square$



---

## References

---

- Abramowitz, M., & Stegun, A. (1972). *Handbook of mathematical functions with formulas, graphs, and mathematical tables*. Wiley, New York.
- Bandi, F. M., & Phillips, P. C. B. (2003). Fully nonparametric estimation of scalar diffusion models. *Econometrica*, 71, 241–283.
- Banon, G. (1978). Nonparametric identification for diffusion processes. *SIAM J. Control Optim.*, 16, 380–395.
- Baraud, Y., Comte, F., & Viennet, G. (2001a). Adaptive estimation in autoregression or  $\beta$ -mixing regression via model selection. *Ann. Statist.*, 29, 839–875.
- Baraud, Y., Comte, F., & Viennet, G. (2001b). Model selection for (auto)-regression with dependent data. *ESAIM Probab. Statist.*, 5, 33–49.
- Barlow, M. T., & Yor, M. (1982). Semimartingale inequalities via the Garsia-Rodemich-Rumsey lemma, and applications to local times. *J. Funct. Anal.*, 49, 198–229.
- Barndorff-Nielsen, O. E. (1978). Hyperbolic distributions and distributions on hyperbolae. *Scand. J. Statist.*, 5, 151–157.
- Barron, A. R., Birgé, L., & Massart, P. (1999). Risk bounds for model selection via penalization. *Probab. Theory Related Fields*, 113, 301–413.
- Beskos, A., Papaspiliopoulos, O., & Roberts, G. O. (2006). Retrospective exact simulation of diffusion sample paths with applications. *Bernoulli*, 12, 1077–1098.
- Beskos, A., & Roberts, G. O. (2005). Exact simulation of diffusions. *Ann. Appl. Probab.*, 15, 2422–2444.
- Bibby, B. M., Jacobsen, M., & Sørensen, M. (2009). Estimating functions for discretely sampled diffusion-type models. In Y. Aït-Sahalia & L. Hansen (Eds.), *Handbook of Financial Econometrics* (pp. 203–268). North Holland, Oxford.
- Bibby, B. M., & Sørensen, M. (1995). Martingale estimation functions for discretely observed diffusion processes. *Bernoulli*, 1, 17–39.
- Birgé, L., & Massart, P. (1998). Minimum contrast estimators on sieves: exponential bounds and rates of convergence. *Bernoulli*, 4, 329–375.
- Comte, F. (2001). Adaptive estimation of the spectrum of a stationary Gaussian sequence. *Bernoulli*, 7, 267–298.



- Comte, F., Genon-Catalot, V., & Rozenholc, Y. (2006). *Nonparametric estimation of a discretely observed integrated diffusion model*. (Tech. Rep.). MAP 5, Mathématiques Appliquées - Paris 5, UMR CNRS 8145.
- Comte, F., Genon-Catalot, V., & Rozenholc, Y. (2007). Penalized nonparametric mean square estimation of the coefficients of diffusion processes. *Bernoulli*, 13, 514–543.
- Comte, F., & Rozenholc, Y. (2002). Adaptive estimation of mean and volatility functions in (auto-)regressive models. *Stochastic Process. Appl.*, 97, 111–145.
- Comte, F., & Rozenholc, Y. (2004). A new algorithm for fixed design regression and denoising. *Ann. Inst. Statist. Math.*, 56, 449–473.
- Dalalyan, A. (2005). Sharp adaptive estimation of the drift function for ergodic diffusions. *Ann. Statist.*, 33, 2507–2528.
- DeVore, R. A., & Lorentz, G. G. (1993). *Constructive Approximation*. Berlin: Springer.
- Florens-Zmirou, D. (1993). On estimating the diffusion coefficient from discrete observations. *J. Appl. Probab.*, 30, 790–804.
- Genon-Catalot, V., Jeantheau, T., & Larédo, C. (2000). Stochastic volatility models as hidden markov models and statistical applications. *Bernoulli*, 6, 1051–1079.
- Genon-Catalot, V., Larédo, C., & Picard, D. (1992). Nonparametric estimation of the diffusion coefficient by wavelet methods. *Scand. J. Statist.*, 19, 319–335.
- Gobet, E., Hoffmann, M., & Reiß, M. (2004). Nonparametric estimation of scalar diffusions based on low frequency data. *Ann. Statist.*, 32, 2223–2253.
- Hoffmann, M. (1999). Adaptive estimation in diffusion processes. *Stochastic Process. Appl.*, 79, 135–163.
- Jacod, J. (2000). Non-parametric kernel estimation of the coefficient of a diffusion. *Scand. J. Statist.*, 27, 83–96.
- Karlin, S., & Taylor, H. M. (1981). *A Second Course in Stochastic Processes*. New York: Academic Press.
- Kessler, M., & Sørensen, M. (1999). Estimating equations based on eigenfunctions for a discretely observed diffusion process. *Bernoulli*, 5, 299–314.
- Kutoyants, Y. A. (2004). *Statistical Inference for Ergodic Diffusion Processes*. London: Springer.
- Pardoux, E., & Veretennikov, A. Y. (2001). On the Poisson equation and diffusion approximation. I. *Ann. Probab.*, 29, 1061–1085.
- Prakasa Rao, B. L. S. (2010). *Statistical Inference for Fractional Diffusion Processes*. Chichester: Wiley.
- Spokoiny, V. G. (2000). Adaptive drift estimation for nonparametric diffusion model. *Ann. Statist.*, 28, 815–836.

## REFERENCES

381

- Tuan, P. D. (1981). Nonparametric estimation of the drift coefficient in the diffusion equation. *Math. Operationsforsch. Statist., Ser. Statistics*, 12, 61–73.