INFERENCE FOR FRACTIONAL STOCHASTIC PROCESSES WITH RANDOM EFFECTS: PARAMETRIC AND NON PARAMETRIC APPROACH

by

Mohamed El Omari

FST, Sultan Moulay Slimane University, 2021

Supervised by

Pr. Hamid El Maroufy

Thesis
Submitted in Partial Fulfillment
of the Requirements for the PhD Degree

Department of Mathematics
Faculty of Science and Techniques
Sultan Moulay Slimane University, Béni-Mellal
January, 2021

THESIS APPROVAL

INFERENCE FOR FRACTIONAL STOCHASTIC PROCESSES WITH RANDOM EFFECTS: PARAMETRIC AND NON PARAMETRIC APPROACH

Dissertation Defended By: **Mohamed El Omari**
Supervised by: **Pr. Hamid El Maroufy**

Dissertation Submitted in Partial Fulfillment of the Requirements for the Degree of Doctor of Philosophy in the field of Statistics

Approved by the committee members:

Mr. Youssef OUKNINE	Professor, Faculty of Sciences Semlalia, Marrakech, Morocco.	**Chair / Reviewer**
Mr. Youssri SLAOUI	Associate Professor, Poitiers University, France.	**Reviewer**
Mr. Abdelkrim MERBOUHA	Professor, Faculty of Sciences and Techniques, Bni- Mellal, Morocco.	**Reviewer**
Mr. Abdessalam BOUAARICH,	Professor, Faculty of Sciences and Techniques, Bni- Mellal, Morocco.	**Examiner**
Mr. Abdelaziz NASROALLAH	Professor, Faculty of Sciences Semlalia, Marrakech, Morocco.	**Examiner**
Mr. Khalifa ES-SEBAIY	Associate Professor, Faculty of Sciences, Kowet University, Kowet.	**Examiner**
Mr. Hamid EL MAROUFY	Professor, Faculty of Sciences and Techniques, Bni- Mellal, Morocco.	**Thesis supervisor**

January 23, 2021

DEDICATION

To my mother FATIMA and my father MOHAMED.

To my wife MARYEM.

ACKNOWLEDGMENTS

Throughout the writing of this dissertation I have received a great deal of support and assistance. I would first like to thank my supervisor, Professor Hamid El Maroufy, for his invaluable assistance and insights leading to bring this dissertation to higher level.

I would like to thank, sincerely, the reviewers Professors Youssef Ouknine, Youssri Slaoui and Abdelkrim Merbouha for their insightful comments and suggestions, which allowed us to improve the quality of the thesis.

I would also like to acknowledge the Professors Abdelaziz Nasroallah, Abdessalam Bouarich and Khalifa Es-Sebaiy for their participation as examiners at the day of dissertation defence. Their fruitful discussions and questions are pushing us to think of some important extensions of this work, in both theoretical and practical point of view.

ABSTRACT

Stochastic differential equation models with random effects are increasingly used in the biomedical fields and have proved to be adequate tools for the study of repeated measurements collected on series of subjects. These models allow the quantification of both between and within subject variation. Performing parametric inference for such models, using discrete (or continuous) time data, is a challenging problem for two reasons: First, the state likelihood is a product of transition densities which are rarely known. Second, the marginalization required to construct this likelihood is an (often multidimensional) integral, which rarely has a closed-form solution.

We provide a class of estimators for Stochastic differential equations (SDE's) with random effects and examine their asymptotic behaviour. We are concerned with SDE's with nonlinear drift and generalized random effects, for which a simulation study is given to highlight the performance of the proposed estimators. We extend the existing results of statistical inference for random effects models to include the SDE's with random effects driven by fractional Brownian motion (fBm). The incorporation of the fBm within our models is of great interest, since it accounts for dependency of increments of the noisy term. This is the case of long-memory phenomena arising in variety of different scientific fields, including hydrology, biology, medicine, economics and traffic network. We consider linear fractional stochastic differential equations with random effects, provide estimators of the common density of random effects, and examine their asymptotic properties. Two types of estimators are considered: kernel density estimators and histogram estimators. Most of our results are illustrated by relevant examples.

ABBREVIATIONS, NOTATIONS AND SYMBOLS

ABBREVIATIONS

RHS Right hand side

LHS Left hand side

REM Random effects models

ODE Ordinary differential equation

SDE Stochastic differential equation

MLE Maximum likelihood estimator

FSDE Fractional stochastic differential equation

O-U Ornstein-Uhlenbeck

Bm Brownian motion

fBm Fractional Brownian motion

r.v Random variable

i.i.d Independent and identically distributed

w.r.t With respect to

BDG Burkholder-Davis-Gundy

GL Globally Lipschitz

LL Locally Lipschitz

Std. dev. Standard deviation

NOTATIONS AND SYMBOLS

$\boldsymbol{a \wedge b}$, $\boldsymbol{a \vee b}$ min$\{a, b\}$, max$\{a, b\}$

$\boldsymbol{\Omega}$, $\boldsymbol{\mathcal{F}}$, $\boldsymbol{\mathcal{B}}$ Sample space, σ-field and Borel σ-field

$\boldsymbol{\sigma(F)}$, $\boldsymbol{(\mathcal{F}_t)}$ σ-field generated by the set F and the filtration

$\mathbb{P}$, $\mathbb{E}$, $\mathbb{V}\boldsymbol{ar}$, $\mathbb{C}\boldsymbol{ov}$ Probability measure, expectation, variance and covariance

$\boldsymbol{\chi_A}$ Indicator function of the set A

supp $\boldsymbol{g}$ support of the function g

$\boldsymbol{L^p}(\mathbb{R})$, $\boldsymbol{L^p([a,b])}$, $\boldsymbol{L^p(\Omega)}$ Spaces of p-integrable functions

$\boldsymbol{L^2}(\mathbb{P})$, $\boldsymbol{L^2(d\nu(\varphi))}$ Spaces of square integrable functions w.r.t the measures $\mathbb{P}$ and ν, respectively

$\boldsymbol{C^\lambda(E)}$ Space of Holder continuous functions on E with exponent λ

$\boldsymbol{C_b^k(E)}$ Space of k-times continuously differentiable functions with bounded derivatives

$\boldsymbol{\overline{B_r(x)}}$ The closed ball centered at x with radius r

$\|\boldsymbol{g}\|$, $\|\boldsymbol{g}\|_{\boldsymbol{p}}$ Euclidian norm and L^p-norm of the function g

$\boldsymbol{g^{\otimes n}}$ The nth power tensor of the function g

$\boldsymbol{\mathcal{N}(\mu, \sigma^2)}$, $\boldsymbol{\mathcal{B}(k, \theta)}$ Normal and Beta distributions

$\boldsymbol{\mathcal{N}^*(\beta, L)}$ Nikol'ski class of functions

$\boldsymbol{o}(\cdot)$, $\boldsymbol{O}(\cdot)$, $\boldsymbol{o_P}(\cdot)$, $\boldsymbol{O_P}(\cdot)$ Usual and stochastic order symbols

$\overset{\mathcal{L}}{\Longrightarrow}$ Convergence in law

$\overset{\mathbb{P}-as}{\Longrightarrow}$ Convergence almost surely (under $\mathbb{P}$)

$\overset{\mathbb{P}}{\Longrightarrow}$ Convergence in probability (under $\mathbb{P}$)

$\longrightarrow$ Simple convergence

$\ll$, $\sim$ Absolute continuity and equivalence of measures

$\lesssim$ Smaller than up to a nonnegative constant

$\frac{d\mu}{d\nu}$ The Radon-Nikodym derivative of the measure μ w.r.t measure ν

$\frac{\partial}{\partial x}$ Partial derivative operator

$< X >$, $< X, Y >$ Quadratic variation process of X and Cross-variation of X and Y

M', M^{-1} The transpose of the matrix M and its inverse

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

CHAPTER 1
INTRODUCTION

Throughout history, interest has lain in understanding and modelling the dynamics of systems evolving through time. Instances include (but are not limited to) the growth of populations, the interactions between certain species, the spread of epidemics and more recently, intra-cellular processes. Initially the dynamics of these systems were captured through the use of ordinary differential equations (ODE's); for example, Kermack and McKendrick [62] describe the spread of a disease through a population using three ODEs. These three ODE's model the changes in the number of individuals who are Susceptible (those who could catch the disease), Infectious (those who have the disease) and Recovered (those who no longer have the disease). This model is known as the SIR model. However, the evolution of these systems is not entirely predictable and is subject to random variation. The deterministic nature of the ODE description is unable to capture this random variation and so has proved to be an unsatisfactory means through which to capture the true dynamics of such systems. Hence an alternative modelling framework is required, which can account for random behaviour. There are two types of randomness which may be considered in the system:

Intrinsic noise, the unexplained variability within the system itself, such as fluctuations in blood pressure, metabolic processes, or varying stress levels. This type of noise can be substantial in biomedical data, because the underlying data generating process is often too complex to be modeled exactly or is not understood well enough. Such internal random fluctuations can be accounted for by including stochasticity in the dynamical model itself. Accounting for this kind of randomness extend the ODE's to stochastic differential equations (SDE's). A system where the introduction of intrinsic noise appears fundamental is the stock market, specifically the pricing of options and shares. Black and Scholes [14] and Merton [78] developed a framework for the fair pricing of options. SDEs consist of both a deterministic and stochastic part, and capture the dynamics of a system through a solution which fluctuates around the deterministic solution. However, it should be noted that the mean of the stochastic solution is not the ODE solution. Some application areas and indicative references where SDEs have been used include finance [23, 12, 18, 59, 106], systems biology [49, 43, 65, 45, 40, 48], population dynamics [47, 53], physics [114, 96, 111], medicine [115, 44, 16], epidemics [24, 3, 51], biology [72], epidemiology [8, 17, 4], genetics [41, 108] and traffic control [76]. The solution of an SDE gives a continuous-time, continuous-valued stochastic process typically referred to as a diffusion process.

Inter-subject fluctuation, that is, the unexplained systematic differences of data dynamics between subjects. Individuals share an overall model structure, or base model, but the values of the model parameters differ between subjects. Parts of that inter-subject variability can generally be captured by including subject-specific covariate information in the model, such as adjustments for gender, age, body weight or treatment group. However, due to the sheer complexity of real systems, a certain amount of unexplained inter-subject variations will remain. The common way to account for them is by imposing random effects on some (or all) parameters. Models that contain both fixed (parameters that are the same across subjects) and random effects are known as mixed-effects models [103, 70, 91]. Accounting only for this kind of randomness leads to ODE's with random effects.

Ordinary differential equations with random effects have frequently been applied to model biomedical data, [97, 58]. Their formulation is intuitive, the random effects capture inter-individual deviations from the population dynamics, and todays computational power renders parameter estimation feasible. Well-known applications of this model framework are pharmacokinetic compartment models [109, 71], which are used to describe the flow of a substance between multiple spatially separated entities (different organs in the human body). Biological systems are, however, incredibly complex. Their variability is driven by the interplay of numerous internal (genetic variations, metabolic fluctuations, etc.) and external factors (stress factors, room temperature, time of day, etc.). The bulk of them can not be measured directly, or can not be included in the model, because it would prohibitively scale up the models complexity. However, ignoring those inadequacies or uncertainties in the model structure lead, if they are substantial, to biased estimates and false inference [30]. In those cases one can achieve a more robust estimation by considering SDE's with random effects. The certainly powerful merging of random effects and SDE's into one single model comes with a considerable challenge for inference based on the data likelihood: its intractability. This now has two sources. First of all, the state likelihood is a product of transition densities which are generally unknown. But even if the transition densities are known, this likelihood has to be marginalized over the distribution of the random effects, because the random effects are practically not observed. Second, the marginalization is an (often multidimensional) integral, which rarely has a closed-form solution. This makes explicit likelihood inference impossible and leaves many research opportunities for finding well-performing numerical or analytical approximation techniques. In fact, numerous ways of tackling this challenge have been explored.

Methods to overcome this difficulty have been proposed, including simulated maximum likelihood estimation [88, 32], closed-form expansion of the transition density [2, 1, 89, 106, 90], exact simulation approaches [11, 10, 9, 102] and Bayesian imputation approaches [35, 39, 98, 49, 107, 66, 100]. The latter method replaces an intractable transition density with a first order Euler-Maruyama approximation, and uses data augmentation to limit the discretisation error incurred by the approximation. Whilst exact algorithms that avoid discretisation error are appealing, they are limited to diffusions which can be transformed to have unit diffusion coefficient, known as reducible diffusions. On the other hand, the Bayesian imputation approach has received much attention in the recent literature due to its wide applicability.

The MLE is usually used to make inference for unknown parameters, since it has a number of desirable properties, such as consistency, asymptotic normality and efficiency. However, as appealing the properties of the MLE may be, it comes with the challenge of an often intractable likelihood. This is the case of nonlinear mixed effects models, where numerical approximations are required (Solving the Kolmogorov forward equations, Gaussian approximations, Hermite expansion, etc.). For theoretical properties of the MLE in the context of mixed effects models, rigorous proofs are less available. the main contribution to our knowledge is due to Nie and Yang[83, 82, 81]. The authors provide the consistency result under several asymptotic frameworks, depending on whether the number of subjects and/or the number of observations per subject goes to infinity. As starting point, we consider SDE's with generalized random effects and study the asymptotic behaviour of the MLE. Then, we provide statistical methods that permit inference for a class of linear FSDE's with random effects. The asymptotic behaviour of our proposed estimators is examined, when the common density of random effects belongs to a class of functions which may not be parametrized. Our results are illustrated by numerical examples.

Outline of thesis :

In the following we outline the subsequent chapters contained within this thesis. Chapter 2 provides preliminary results about stochastic calculus: Itô Calculus and Wiener Integrals w.r.t the fBm. Chapter 3 is devoted to the literature review about random-effects models, which are the main subject of our thesis. A general model gathering specific cases treated separately in Chapter 4 & 5 is presented. This chapter contains also some auxiliary results (of great interest) used systematically in Chapter 4 to establish the asymptotic behaviour of our estimators. In Chapter 4 we discuss the problem of parametric estimation for diffusion processes with generalized random effects. This chapter is split into two sections. Section 4.1 provides a class of estimators of the population parameters, which satisfy the consistency and the asymptotic normality; while Section 4.2 is devoted to weakening the imposed assumptions. An expasion of the likelihood function is established for this purpose.

A class of linear FSDE's with random effects is presented in Chapter 5. We distinguish two problems of inference: parametric estimation in Section 5.2 and non parametric estimation in Sections 5.3-5.4; while Section 5.1 presents the general model considered within this chapter. The asymptotic behaviour of our estimators is studied in detail for both types of inference. We initially discuss the challenging problem of constructing the likelihood. The main obstacle is the non Markovian and nonsemimartingale nature of the fBm. Thus the classical techniques used in Chapter 4 are not applicable. However, we consider a simple model that is of great interest in finance, and derive explicit estimators of random effects parameters. Our results and numerical simulations for this model are given in Section 5.2. In Section 5.3 we provide estimators of the commun density f of random effects, for which we study the L^p-risk ($p = 1$, or 2). Two types of estimators are considered: kernel density estimators and histogram estimators. Our results are discussed and implemented for an O-U process. Section 5.4 is devoted to the problem of non parametric estimation for a linear FSDE with small diffusion. Both L^2-risk and pointwise-risk of our estimators are examined and enhanced by numerical examples. Our examples are implemented for different drift terms and different density functions of random effects. Finally, concluding remarks and perspectives are given in Chapter 6. For the clarity of exposure some results and tedious computations are gathered in Appendix.

CHAPTER 2
PRELIMINARIES

We assume the reader is familiar with basic concepts of probability theory and theory of statistics. Such background material can be found in [105, 101, 113]. This chapter provides auxiliary results about Itô calculus and stochastic calculus for fBm that are employed throughout this thesis. Most are well known and stated without proof, as they can be found in standard literature, such as [73, 69, 33] and [94, 84, 80], respectively.

2.1 ITÔ CALCULUS

2.1.1 The Kolmogorov-Chentsov Theorem

Theorem 2.1.1. ***(Karatzas,[60, p. 53])*** *Suppose that a process $X = (X_t \,:\, 0 \leq t \leq T)$ on a probability space $(\Omega, \mathcal{F}, \mathbb{P})$ satisfies the condition*

$$\mathbb{E}\left|X_t - X_s\right|^{\alpha} \;\leq\; C\left|t-s\right|^{1+\beta}, \quad 0 \leq s,t \leq T, \tag{2.1}$$

for some nonnegative constants α, β and C. Then there exists a continuous modification $\widetilde{X} = \left(\widetilde{X}_t \,:\, 0 \leq t \leq T\right)$ of X, which is locally Hölder-continuous with exponent γ for every $\gamma \in (0, \beta/\alpha)$, i.e.,

$$\mathbb{P}\left(\omega \;:\; \sup_{\substack{0\leq|t-s|\leq h(\omega)\\ s,t\in[0,T]}} \frac{\left|\widetilde{X}_t - \widetilde{X}_s\right|}{\left|t-s\right|^{\gamma}} \leq \delta\right) \;=\; 1, \tag{2.2}$$

where $h(\omega)$ is $\mathbb{P}$-a.s nonnegative r.v and $\delta > 0$ is an appropriate constant.

A random field is a collection of r.v's $X = (X_t \,:\, t \in A)$, where A is partially ordered. An example of random field is $X = \left(X_t \,:\, t \in [0,T]^d\right)$, with $d \geq 2$ is an integer. In this case the Kolmogorov-Chentsov criterion [20, p. 36] is given by

$$\mathbb{E}\left|X_t - X_s\right|^{\alpha} \;\leq\; C\left\|t-s\right\|^{d+\beta}, \quad 0 \leq s,t \leq T. \tag{2.3}$$

2.1.2 Brownian Motion Processes

In 1828 the Scottish botanist Robert Brown discovered Brownian motion after examining pollen from a plant suspended in water under the lens of a microscope (see, [15]). He noted that minute particles ejected from the pollen grain displayed a continuous irregular motion. In 1900 the French mathematician Louis Bachelier [7] considered Brownian motion as a model for stock, mathematically defining Brownian motion in the process. The governing laws of Brownian motion were established by Albert Einstein [34]. Norbert Wiener [116] proved the existence (and provided the construction) of Brownian motion, and it is for this reason that Brownian motion is also referred to as the Wiener process. The univariate stochastic process $(W_t \ : \ t \geq 0)$ is defined to be Brownian motion if $W_t \in \mathbb{R}$ depends continuously on t and the following conditions hold

i) $\mathbb{P}\,(W_0 = 0) = 1$;

ii) W is a process with stationary independent increments, i.e., for all times $0 \leq t_0 < t_1 < t_2$, the r.v's $(W_{t_2} - W_{t_1})$ and $(W_{t_1} - W_{t_0})$ are independent;

iii) increments $W_{t_2} - W_{t_1}$ have a Gaussian normal distribution with

$$\mathbb{E}(W_{t_2} - W_{t_1}) \ = \ 0 \ \text{ and } \ \mathbb{V}ar\,(W_{t_2} - W_{t_1}) = \sigma^2\,|t_2 - t_1|\,. \tag{2.4}$$

In the case $\sigma^2 = 1$, the process W is often called the standard Brownian motion process. The existence of such a process on (fairly 'rich') probability spaces may be established in a constructive way (see, [73, Theorem 1.13]).

Remark. The Brownian motion satisfies many properties, such as, the law of the iterated logarithm and the Hölder condition of Lévy (e.g., [73, p. 32]). We focus here on properties which will allow us to construct stochastic integrals of the form $\displaystyle\int Y_s(\omega)dW_s$, where $Y_t(\omega)$ belongs to some class of random functions. This kind of integrals cannot be defined as Lebesgue-Stieltjes or Riemann-Stieltjes integrals, since realizations of a Bm have unbounded variation in any arbitrary small interval of time. However, the following result shows that Bm trajectories have some properties which in some sense are analogous to bounded variation.

Proposition 2.1.2. *Let $0 = t_0^{(n)} < t_1^{(n)} < \cdots < t_n^{(n)} = t$ be a subdivision of the interval $[0,t]$, with $\pi^{(n)} = \max_i \left\{ t_{i+1}^{(n)} - t_i^{(n)} \right\} \to 0$, as $n \to \infty$. Then*

$$< W >_t \ := \ \underset{n\to\infty}{Lim} \sum_{i=1}^{2^n} \left(W\left(\frac{it}{2^n}\right) - W\left(\frac{(i-1)t}{2^n}\right)\right)^2 \tag{2.5}$$

exists $\mathbb{P}$-a.s and $< W >_t = t$. The limit $< W >_t$ is the value of the quadratic variation of the Brownian motion W at time t.

We will later define the quadratic variation of a right-continuous martingale. The definition will be given as a result of Doob-Meyer decomposition (for more details about this topic, we refer the reader to [60, p. 24]).

Corollary 2.1.3. *The Brownian motion is not of finite variation, i.e.,*

$$\underset{n\to\infty}{Lim}\sum_{i=1}^{2^n}\left|W\left(\frac{it}{2^n}\right)-W\left(\frac{(i-1)t}{2^n}\right)\right| \;=\; \infty, \quad \mathbb{P}\text{-}a.s \tag{2.6}$$

2.1.3 Martingales and Related Processes

Martingales are a very important subject in their own right as well as by their relationship with analysis. Their kinship to Bm and their contribution to the construction of stochastic integrals will make them one of our foremost tools. This section describes some of their basic properties.

Definition. A real-valued process $M = (M_t \;:\; 0 \leq t \leq T)$ adapted to $(\mathcal{F}_t)_t$ is a supermartingale (w.r.t $(\mathcal{F}_t)_t$) if

i) For any $t \in [0,T]$, $\mathbb{E}\left|M_t\right| < \infty$;

ii) For any pair s,t such that $s < t$, we have $\mathbb{E}\left(M_t|\mathcal{F}_s\right) \leq M_s$, $\mathbb{P}$-a.s.

A process M such that $-M$ is supermartingale is called submartingale and a process which is both a sub and a supermartingale is a martingale.

In other words, a martingale is an adapted family of integrable r.v's M such that

$$\int_A M_s d\mathbb{P} \;=\; \int_A M_t d\mathbb{P}, \quad \text{for every pair } s,t \text{ with } s < t \text{ and } A \in \mathcal{F}_s. \tag{2.7}$$

In particular, $\mathbb{E}(M_s) = \mathbb{E}(M_t)$ for all $t \geq s$. A simple example we give here is the process $M = (E\ \ (X|\mathcal{F}_t) \;:\; 0 \leq t \leq T)$, where X is a r.v with $\mathbb{E}(|X|) < \infty$.

Proposition 2.1.4. *Let W be a Brownian motion. Then the following processes are martingales w.r.t the filtration generated by W*

i) W_t itself;

ii) The process $W_t^2 - t$;

iii) The nonnegative process $M_t^\alpha \;=\; \exp\left(\alpha W_t - \frac{\alpha^2}{2}t\right)$, $\alpha \in \mathbb{R}$.

Let $\mathcal{M}_2$ be the set of right-continuous martingales such that $\mathbb{E}M^2 < \infty$ and $M_0 = 0$. Let $\mathcal{S}_a$ denotes the set of stopping times (see, [60, p. 6]) bounded almost surely by a given number $a > 0$, we say that M is of class **DL** if the familly $\{M_T\}_{T \in \mathcal{S}_a}$ is uniformly integrable. We define **the quadratic variation** of M as the unique (up to indistinguishability) adapted, natural nondecreasing process $< M >$, for wich $< M >_0 = 0$ a.s and $M^2 - < M >$ is a martingale.

Remark. It is also convenient to define the quadratic variation $< M >$ of a **continuous martingale** $M = (M_t, \mathcal{F}_t)$ at time $t > 0$ as the limit in probability of the following sums

$$V_t^{(2)} := \sum_{k=1}^{m} \left| M_{t_k} - M_{t_{k-1}} \right|^2, \quad \text{as } \|\pi\| \to 0, \tag{2.8}$$

where $\|\pi\| = \max\limits_{1 \le k \le m} \left| t_k^{(m)} - t_{k-1}^{(m)} \right|$ and $0 = t_0^{(m)} < t_1^{(m)} < \cdots < t_m^{(m)} = t$ is a subdivision of the interval $[0, t]$.

Let $Z_t^{(1)}$ and $Z_t^{(2)}$ be two martingales, we define the bracket of $Z_t^{(1)}$ and $Z_t^{(2)}$ by

$$< Z^{(1)}, Z^{(2)} >_t = \frac{1}{4} \left\{ < Z^{(1)} + Z^{(2)} >_t - < Z^{(1)} - Z^{(2)} >_t \right\}. \tag{2.9}$$

Obviously, the process $< Z^{(1)}, Z^{(2)} >_t$ is the limit in probability of $\sum_{k=1}^{m} \left(Z_{t_k}^{(1)} - Z_{t_{k-1}}^{(1)} \right) \left(Z_{t_k}^{(2)} - Z_{t_{k-1}}^{(2)} \right)$.

The following results provide a strong bridge between the Brownian motions processes and the class of contintuous square integrable martingales. Thus, it is possible to extend the stochastic integrals w.r.t Bm to integrals with martingale as integrator process.

Theorem 2.1.5. ***(P. Lévy [73, p. 85]).*** *Let $X = (X_t, \mathcal{F}_t)_{t \ge 0}$ be a continuous, adapted process in $\mathbb{R}$ such that $M_t = X_t - X_0$ is a continuous martingale w.r.t $(\mathcal{F}_t)$ and $< X >_t = t$ for all $t \ge 0$. Then X is a Brownian motion.*

Theorem 2.1.6. *Let $W = (W_t, \mathcal{F}_t)$ be a Brownian motion. Suppose that $M = (M_t, \mathcal{F}_t)$ is a continuous square integrable martingale (w.r.t $\mathbb{P}$). then there exists a unique process $Y_t(\omega)$ such that $\mathbb{E} \int_0^t Y_s^2 ds < \infty$ and $M_t(\omega) = \mathbb{E}(M_0) + \int_0^t Y_s(\omega) dW_s$, $\mathbb{P}$-a.s for all $t \ge 0$, where $\int_0^t Y_s(\omega) dW_s$ is an Itô integral defined in the next subsection.*

Theorem 2.1.7. ***(Burkholder-Davis-Gundy [60, p. 166]).*** *Let M be a continuous martingale. For any $m > 0$, there exist universal nonnegative constants c_m, C_m (depending only on m) such that*

$$c_m \mathbb{E}\left(< M >_T^m\right) \;\le\; \mathbb{E}\left(\sup_{0\le t\le T} M_t^{2m}\right) \;\le\; C_m \mathbb{E}\left(< M >_T^m\right), \tag{2.10}$$

provided that $\mathbb{E}\left(< M >_T^m\right) < \infty$.

2.1.4 The Itô Integral

Let us now introduce the class of random functions $Y_t(\omega)$, for which the stochastic integral $I_t(Y) := \int_0^t Y_s(\omega) dW_s$ is well defined.

Definition. The measurable (w.r.t a pair of variables (t, ω)) function $Y = Y_t(\omega)$, $t \ge 0$, $\omega \in \Omega$ is called nonanticipative w.r.t the family $F = (\mathcal{F}_t), t \ge 0$, if for each t, it is $\mathcal{F}_t$-measurable.

Definition. The nonanticipative function $Y = Y_t(\omega)$ is said to be of class $\mathcal{P}_T$ if

$$\mathbb{P}\left(\int_0^T Y_t(\omega)^2 dt < \infty\right) = 1. \tag{2.11}$$

Definition. The nonanticipative function $Y = Y_t(\omega)$ is said to be of class $\mathcal{V}_T$ if

$$\mathbb{E}\int_0^T Y_t(\omega)^2 dt < \infty. \tag{2.12}$$

The nonanticipative functions defined above are measurable random processes, adapted to the family $(\mathcal{F}_t)$. Obviously, for any $T \ge 0$, $\mathcal{V}_T \subseteq \mathcal{P}_T$. By analogy with the conventional integration theory it is natural to define first the stochastic integral $I_t(Y)$ for a certain set of "elementary "functions Y. this set has to be sufficiently "rich"; so that $I_t(Y)$ can be easily constructed on this set, and any function from $\mathcal{P}_T$ or $\mathcal{V}_T$ can be approximated by functions of this set. Such set of "elementary "functions is denoted by $\mathcal{E}_T$ and consists of simple functions introduced in the definition below.

Definition. The function $e = e(t, \omega)$, $0 \le t \le T$ is called simple if there exist a finite subdivision $0 = t_0^{(n)} < t_1^{(n)} < \cdots < t_n^{(n)} = T$ of the interval $[0, T]$, random variables $\alpha, \alpha_0, \cdots, \alpha_{n-1}$, where α is $\mathcal{F}_0$-measurable and α_i are $\mathcal{F}_{t_i}$-measurable, $i = 1, \cdots, n-1$ such that

$$e(t, \omega) = \alpha \chi_{\{0\}}(t) + \sum_{i=0}^{n-1} \alpha_i \chi_{(t_i, t_{i+1}]}(t),$$

where χ_A denotes the chracteristic function of the set A and $e(t, \omega) \in \mathcal{V}_T$.

With these arguments, the integral $I_t(e)$ is given by

$$I_t(e) = \sum_{\substack{0 \le i \le m \\ t_{m+1} < t}} \alpha_i \left(W_{t_{i+1}} - W_{t_i} \right) + \alpha_{m+1} \left(W_t - W_{t_{m+1}} \right). \tag{2.13}$$

Instead of the sums given in (2.13), we shall use the following notation $I_t(e) := \int_0^t e(s,\omega) dW_s$. The integral $\int_s^t e(u,\omega) dW_u$ will be understood as $\int_s^t e(u,\omega) dW_u = \int_0^t e(u,\omega)\chi_{(u>s)} dW_u$. The main properties of the stochastic integral (Itô integral) $I_t(\cdot)$ for simple functions are summarized below

1. **Linearity :** For all $a, b \in \mathbb{R}$ and $e_1, e_2 \in \mathcal{E}_T$,

$$I_t(ae_1 + be_2) = aI_t(e_1) + bI_t(e_2);$$

2. $I_t(e)$ is a continuous function over t, $t \in [0,T]$;

3. **Martingale property :** For all $s \le t$, we have

$$\mathbb{E}\left(I_t(e)|\mathcal{F}_s\right) = I_s(e);\ \text{In particular } \mathbb{E}\left(\int_s^t e(u,\omega) dW_u\right) = 0, \quad \text{for all } s \le t;$$

4. **Generalized Itô isometry :** For all s, t, we have

$$\mathbb{E}\left(\int_0^s e_1(u,\omega) dW_u \int_0^t e_2(u,\omega) dW_u\right) = \mathbb{E}\int_0^{s\wedge t} e_1(u,\omega) e_2(u,\omega) du$$

For the case where $Y_t(\omega) \in \mathcal{V}_T$ (or $Y_t(\omega) \in \mathcal{P}_T$), the stochastic integral (Itô integral) $I_t(Y)$ is defined as the limit of $\int_0^t Y_u(\omega)^{(n)} dW_u,\ Y_t(\omega)^{(n)} \in \mathcal{E}_T$ in L^2-sense (with probability one), respectively (see, [73]). For simple functions $Y_t(\omega)$, the Itô integral $I_t(Y)$ can be verified directly using (2.13). As an illustration, consider the case $Y = c \in \mathbb{R}$. We have

$$\begin{aligned} \int_0^t c dW_s &= \underset{n\to\infty}{\text{l.i.m}} \sum_{i=1}^n c \left(W_{t_i}^{(n)} - W_{t_{i-1}}^{(n)} \right) \\ &= \underset{n\to\infty}{\text{l.i.m}}\ c\left(W_t - W_0\right) = cW_t, \end{aligned}$$

where $0 = t_0^{(n)} < t_1^{(n)} < \cdots < t_n^{(n)} = t$ is a subdivision of the interval $[0,t]$ with $\max_i \left| t_i^{(n)} - t_{i-1}^{(n)} \right| \to 0$ as $n \to \infty$ and l.i.m denotes the limit with probability one. Another interesting example is the case where Y is a Brownian motion W itself. straightforward computations lead to

$$\begin{aligned}\int_0^t W_s dW_s &= \underset{n\to\infty}{\text{l.i.m}} \sum_{i=1}^n W_{t_{i-1}}^{(n)} \left(W_{t_i}^{(n)} - W_{t_{i-1}}^{(n)} \right) \\ &= \underset{n\to\infty}{\text{l.i.m}} \left[-\frac{1}{2} \sum_{i=1}^n \left(W_{t_i}^{(n)} - W_{t_{i-1}}^{(n)} \right)^2 + \frac{1}{2} \sum_{i=1}^n \left((W_{t_i}^{(n)})^2 - (W_{t_{i-1}}^{(n)})^2 \right) \right] \\ &= -\frac{1}{2} t + \frac{1}{2} W_t^2. \end{aligned}$$

Remark.

1. If $Y \in \mathcal{V}_T$, then the Itô integral $I_t(Y)$ is continuous square integrable martingale with quadratic variation
$$< I_.(Y) >_t = \int_0^t Y_s(\omega)^2 ds, \tag{2.14}$$
but not necessarily Gaussian;

2. The process $I_t(Y),\ 0 \le t \le T$ in the case $Y \in \mathcal{P}_T$, is, generally speaking, not a martingale.

Theorem 2.1.8. ***(Integration by parts formula [99, p. 59]).*** *If $Z^{(1)}$ and $Z^{(2)}$ are two continuous martingales. Then*

$$d\left(Z_t^{(1)} Z_t^{(2)}\right) = Z_t^{(1)} dZ_t^{(2)} + Z_t^{(2)} dZ_t^{(1)} + d < Z^{(1)}, Z^{(2)} >_t . \tag{2.15}$$

In particular, $d\left([Z_t^{(1)}]^2\right) = 2Z_t^{(1)} dZ_t^{(1)} + d < Z^{(1)} >_t$.

2.1.5 The Itô Formula

Throughout this section, we consider the diffusion processes of the form

$$\begin{aligned} \xi_t &= \xi_0 + \int_0^t A(s,\xi_s)ds + \int_0^t B(s,\xi_s)dW_s, \quad 0 \le t \le T, \\ \text{where} \qquad & \mathbb{P}\left(\int_0^T |A(s,\xi_s)| + B^2(s,\xi_s)ds \right) = 1. \end{aligned}$$

In differential form, we write

$$d\xi_t = A(t,\xi_t)dt + B(t,\xi_t)dW_t, \quad 0 \le t \le T. \tag{2.16}$$

Theorem 2.1.9. ***(Itô formula [73, p. 124]).*** *Let $f(t,x)$ be a continuous function and have partial derivatives $f'_t(t,x)$, $f'_x(t,x)$ and $f''_x x(t,x)$. Assume the random process $\xi = (\xi_t, \mathcal{F}_t)$, $0 \le t \le T$ has the stochastic differential form given by (2.16). Then*

$$df(t,\xi_t) = f'_t(t,\xi_t)dt + f'_x(t,\xi_t)d\xi_t + \frac{1}{2}f''_x x(t,\xi_t)B^2(t,\xi_t)dt \tag{2.17}$$

The Itô formula is powerful tool in stochastic calculus. It is usually applied to solve SDE's or compute integrals. For example, one can compute the integral $\int_0^t W_s^2 dW_s$ by applying Itô formula to the function $f(t,W_t) = W_t^3$. In fact $d\left(W_t^3\right) = 3W_t^2 dW_t + 3W_t dt$. Thus $\int_0^t W_s^2 dW_s = \frac{1}{3}W_t^3 - \int_0^t W_s ds$.

Consider the SDE

$$d\xi_t = \alpha\xi_t dW_t,\ 0 \le t \le T,\ \alpha \in \mathbb{R},\ \xi_0, \tag{2.18}$$

where ξ_0 and $(W_t, \mathcal{F}_t)$ are $\mathcal{F}_0$-measurable r.v and Bm, respectively. It is easy to see that $X_t = \xi_0 \exp\left(\alpha W_t - \frac{1}{2}\alpha^2 t\right)$, $0 \le t \le T$ solves (2.18). To prove uniqueness, Let $Y = (Y_t, \mathcal{F}_t)$ be another solution to (2.18) starting at $Y_0 = \xi_0$ and set $Z_t = \exp\left(-\alpha W_t + \frac{1}{2}\alpha^2 t\right)$. By the integration by parts formula (2.15), we have

$$\begin{aligned} d\left(Z_t Y_t\right) &= Z_t dY_t + Y_t dZ_t + d < Z, Y >_t \\ &= Z_t\left(\alpha Y_t dW_t\right) + Y_t\left[\frac{1}{2}\alpha^2 Z_t dt - \alpha Z_t dW_t + \frac{1}{2}\alpha^2 Z_t dt\right] - \alpha^2 Z_t Y_t dt \\ &= 0. \end{aligned}$$

Thus $Z_t Y_t$ is constant, and so $Y_t = Z_0 Y_0 / Z_t = X_t$.

2.1.6 Strong Solutions of Stochastic Differential Equations

In this section, we are interested in homogeneous diffusion processes defined as solutions to the SDE

$$d\xi_t = A(\xi_t)dt + B(\xi_t)dW_t,\ \xi_0,\quad 0 \le t \le T, \tag{2.19}$$

where $A(x)$, $B(x)$ are nonrandom functions and are called trend coefficient and diffusion coefficient, respectively. The equation (2.19) should be understood as a short version of the following integral equation

$$\xi_t = \xi_0 + \int_0^t A(\xi_s)ds + \int_0^t B(\xi_s)dW_s,\ 0 \le t \le T,$$

provided that $$\mathbb{P}\left(\int_0^t (|A(\xi_s)| + B^2(\xi_s))ds < \infty\right) = 1.$$

Let us introduce a family of σ-algebras $\mathcal{F}_t^{\xi_0,W} := \sigma\{\xi_0,\ W_s,\ 0 \le s \le t\}$, $0 \le t \le T$ generated by the initial value ξ_0 and by the given Wiener process up to time t.

Definition. Equation (2.19) has a strong solution $\xi = (\xi_t\ :\ 0 \le t \le T)$ on the given probability space $(\Omega, \mathcal{F}, \mathbb{P})$ w.r.t the fixed Wiener process $W = (W_t\ :\ 0 \le t \le T)$ and initial condition ξ_0 if the process ξ satisfies the equality (2.19), has continuous sample paths and ξ_t is $\mathcal{F}_t^{\xi_0,W}$-measurable for all $t \in [0, T]$.

We say that the SDE (2.19) has unique strong solution if for any two solutions $\xi^{(1)} = \left(\xi_t^{(1)}\ :\ 0 \le t \le T\right)$ and $\xi^{(2)} = \left(\xi_t^{(2)}\ :\ 0 \le t \le T\right)$ the following equality holds true

$$\mathbb{P}\left(\sup_{0\le t\le T}\left|\xi_t^{(1)} - \xi_t^{(2)}\right| > 0\right) = 0.$$

There are two type of solutions to (2.19). We focus here on strong solutions, for which many conditions insuring their existence and uniqueness are given below.

GL. (Globally Lipschitz condition). There exists a constant $L > 0$ such that

$$|A(x) - A(y)| + |B(x) - B(y)| \le L\,|x - y|\,, \quad \text{for all } x, y \in \mathbb{R}$$

Theorem 2.1.10. ***(Kutoyants [69, p. 25])*** *Let the condition* ***GL*** *be fulfilled and $\mathbb{P}(|\xi_0| < \infty) = 1$. Then the equation (2.19) has a unique (strong) solution $(\xi_t\ :\ 0 \le t \le T)$, continuous $\mathbb{P}$-a.s. If moreover $\mathbb{E}\,|\xi_0|^{2m} < \infty$, then $\mathbb{E}\,|\xi_t|^{2m} \le \left(1 + \mathbb{E}\,|\xi_0|^{2m}\right)e^{c_m t} - 1$, where c_m is some nonnegative constant.*

LL. (Locally Lipschitz condition). For any $M < \infty$ and $|x|, |y| \le M$ there exists a constant $L_M > 0$ such that $|A(x) - A(y)| + |B(x) - B(y)| \le L_M\,|x - y|$ and $2xA(x) + B^2(x) \le L\left(1 + x^2\right)$, for some constant $L > 0$.

Of course the condition **LL** is less restrictive and is fulfilled for many examples, such as the Double-Well Potential diffusion, i.e.,

$$dX_t = \left(v_1 X_t - v_2 X_t^3\right)dt + v_3 dW_t,\ X_0,\ 0 \le t \le T,$$

where $v_i,\ i = 1, 2, 3$ are nonnegative parameters in compact space.

Theorem 2.1.11. ***(Kutoyants [69, p. 26])*** *Let the condition* ***LL*** *be fulfilled and $\mathbb{P}(|\xi_0| < \infty) = 1$. Then the equation (2.19) has a unique (strong) solution $(\xi_t\ :\ 0 \le t \le T)$, continuous $\mathbb{P}$-a.s.*

For multidimensional case, where $x \in \mathbb{R}^n$, $A(x) = (A_1(x), A_2(x), \cdots, A_n(x))'$ and $B(x) = (B_{i,j}(x))_{1\le i,j\le n}$, the locally Lipschitz condition **LL** is given by

(i) For any $M < \infty$ and $\|x\|, \|y\| \le M$, there exists a constant $L_M > 0$ such that

$$\sum_i |A_i(x) - A_i(y)| + \sum_{i,j} |B_{i,j}(x) - B_{i,j}(y)| \le L_M \|x - y\|;$$

(ii) There exists $L > 0$ such that

$$\sum_i (2x_i A_i(x) + b_{ii}(x)) \le L\left(1 + \|x\|^2\right), \text{ where } b(x) = B(x)B(x)'.$$

2.1.7 Nonnegative Supermartingales and Girshanov Theorem

Let $(\Omega, \mathcal{F}, \mathbb{P})$ be a complete probability space and let $(\mathcal{F}_t)$, $0 \le t \le T$ be a nondecreasing of sub-σ-algebras of $\mathcal{F}$, augmented by sets from $\mathcal{F}$ of probability zero. Let $W = (W_t, \mathcal{F}_t)$ be a Wiener process and let $\gamma = (\gamma_t, \mathcal{F}_t)$ be a random process with $\mathbb{P}\left(\int_0^T \gamma_s^2 ds < \infty\right) = 1$. The investigation of the problem of absolute continuity of measures of diffusion processes w.r.t a Wiener measure relies on the Girshanov theorem. In establishing Girshanov theorem (which provides a powerful tool for change of measures), an essential role is played by nonnegative continuous ($\mathbb{P}$-a.s) processes $\kappa = (\kappa_t, \mathcal{F}_t)$, $0 \le t \le T$ permitting the representation

$$\kappa_t = 1 + \int_0^t \gamma_s dW_s. \tag{2.20}$$

Proposition 2.1.12. *Let $\gamma = (\gamma_t, \mathcal{F}_t)$, $t \le T$ satisfy (2.20) and $\kappa_t \ge 0$, ($\mathbb{P}$-a.s), $0 \le t \le T$. Then the process $\kappa = (\kappa_t, \mathcal{F}_t)$ is a nonnegative supermartingale.*

An important particular case of nonnegative continuous ($\mathbb{P}$-a.s) supermartingales permitting representation given by (2.20) is represented by processes ψ of the form

$$\psi_t = \exp\left(\int_0^t \alpha_s dW_s - \frac{1}{2}\int_0^t \alpha_s^2 ds\right), \quad t \le T, \tag{2.21}$$

where the process $\alpha = (\alpha_t, \mathcal{F}_t)$ is such that $\mathbb{P}\left(\int_0^T \alpha_s^2 < \infty\right) = 1$. By virtue of Itô formula it follows that ψ admits the representation (2.20), with $\gamma_t = \alpha_t \psi_t$. The following theorems provide conditions for which supermartingales of the form (2.20) are true martingales.

Theorem 2.1.13. ***(Liptser [73, p. 228])*** *Let $\kappa = (\kappa_t, \mathcal{F}_t)$, $t \le T$ be a supermartingale of the form (2.20) such that $\mathbb{E}(\kappa_T) = 1$. Then κ is a true martingale.*

Theorem 2.1.14. ***(Novikov condition [73, p. 229]).*** *Let $\alpha = (\alpha_t, \mathcal{F}_t)$, $t \le T$ be a random process such that*

$$\mathbb{E}\exp\left(\frac{1}{2}\int_0^T \alpha_s^2 ds\right) < \infty.$$

Then the process ψ given in (2.21) is a true martingale, and in particular $\mathbb{E}(\psi_t) = 1,\ t \le T$.

Let $\kappa = (\kappa_t, \mathcal{F}_t)$, $t \le T$ be a nonnegative continuous supermartingale with $\kappa_t = 1 + \int_0^t \gamma_s dW_s$ and $\mathbb{E}(\kappa_T) = 1$. The process κ is a nonnegative martingale, and on the measurable space $(\Omega, \mathcal{F}_T)$, we can define the measure $\widetilde{\mathbb{P}}$ by $d\widetilde{\mathbb{P}} = \kappa_T(\omega)d\mathbb{P}$. We denote by κ_s^+ the process defined by

$$\kappa_s^+ = \begin{cases} 1/\kappa_s\ , & \kappa_s > 0; \\ 0\ , & \kappa_s = 0. \end{cases}$$

Theorem 2.1.15. ***(Girshanov theorem [73, p. 238]).*** *On the probability space $\left(\Omega, \mathcal{F}_T, \widetilde{\mathbb{P}}\right)$ the random process $\widetilde{W} = (\widetilde{W}_t, \mathcal{F}_t)$, $t \le T$, with*

$$\widetilde{W}_t = W_t - \int_0^t \kappa_s^+ \gamma_s ds$$

is a Wiener process under the measure $\widetilde{\mathbb{P}}$.

2.1.8 Absolute Continuity of Measures Corresponding to Diffusion Processes

The following theorem is of great interest for the construction of likelihood functions. This theorem will be systematically used in Chapter 4. As an illustration the problem of parametric estimation for the ergodic O-U process is investegated below.

Theorem 2.1.16. ***(Liptser [73, p. 294])*** *Let $\xi = (\xi_t, \mathcal{F}_t)$ and $\eta = (\eta_t, \mathcal{F}_t)$, $0 \le t \le T$ be two processes of the diffusion type with*

$$\begin{cases} d\xi_t &= A_1(\xi_t)dt + B(\xi_t)dW_t, \quad \xi_0 = \eta_0, \\ d\eta_t &= A_2(\eta_t)dt + B(\eta_t)dW_t, \end{cases} \tag{2.22}$$

where ξ_0 is $\mathcal{F}_0$-measurable r.v and $W = (W_t, \mathcal{F}_t)$ is a Wiener process. Let the following assumptions be fulfilled.

I) *The functions $A_i(x)$, $i=1,2$ and $B(x)$ satisfy any conditiondsproviding the existence and uniqueness of a strong solution to the system (2.1.16);*

II) *For any t, $0 \le t \le T$, the equation $B(\xi_t)\alpha_t(\omega) = A_1(\xi_t) - A_2(\xi_t)$ has (w.r.t $\alpha_t(\omega)$) $\mathbb{P}$-a.s solution;*

III) *The following equality holds*

$$\begin{aligned}
&\mathbb{P}\left(\int_0^T (B^+(\xi_s))^2 \left(A_1^2(\xi_s) + A_2^2(\xi_s)\right) ds < \infty\right) \\
= \; &\mathbb{P}\left(\int_0^T (B^+(\eta_s))^2 \left(A_1^2(\eta_s) + A_2^2(\eta_s)\right) ds < \infty\right) = 1.
\end{aligned}$$

Then $\mu_\xi \sim \mu_\eta$ and the density $\dfrac{d\mu_\xi}{d\mu_\eta}(\eta)$ is given by

$$\begin{aligned}
\frac{d\mu_\xi}{d\mu_\eta}(\eta) \;=\; \exp\Bigg[&\int_0^T (B^+(\eta_s))^2 \left(A_1(\eta_s) - A_2(\eta_s)\right) d\eta_s \\
&-\frac{1}{2}\int_0^T (B^+(\eta_s))^2 \left(A_1^2(\eta_s) - A_2^2(\eta_s)\right) ds\Bigg]. \qquad (2.23)
\end{aligned}$$

Example. Let us consider the problem of parameter estimation by the observations $X = (X_t \;:\; 0 \le t \le T)$ of the O-U process

$$dX_t = -\lambda X_t dt + dW_t, \; X_0, \; 0 \le t \le T,$$

where the unknown parameter belongs to $[\underline{\lambda}, \overline{\lambda}] \subset (0,\infty)$. Let λ_0 be true value of the parameter λ. Applying Theorem 2.1.16 to the processes $\xi_t = X_t^{(\lambda)}$ and $\eta_t = X_t^{(\lambda_0)}$, we obtain (by formula (2.23)) as likelihood

$$L_T(\lambda) \;=\; \exp\left[-(\lambda-\lambda_0)\int_0^T X_t^{(\lambda_0)} dX_t^{(\lambda_0)} - \frac{1}{2}(\lambda^2-\lambda_0^2)\int_0^T (X_t^{(\lambda_0)})^2 dt\right].$$

Thus the MLE is given by $\widehat{\lambda}_T = -\int_0^T X_t dX_t / \int_0^T X_t^2 dt$, and

$$\begin{aligned}
\sqrt{T}\left(\widehat{\lambda}_T - \lambda_0\right) \;&=\; \left(\frac{1}{T}\int_0^T X_t^2 dt\right)^{-1}\left(\frac{1}{\sqrt{T}}\int_0^T X_t dW_t\right) \\
&\stackrel{\mathcal{L}}{\Longrightarrow} \; \mathcal{N}(0, 2\lambda_0).
\end{aligned}$$

2.2 STOCHASTIC INTEGRATION WITH RESPECT TO FBM AND RELATED TOPICS

2.2.1 Intrinsic Properties of the Fractional Brownian Motion

In this subsection we review the main properties that make fractional Brownian motion interesting for many applications in different fields. The main references for this subsection are [74, 85, 99, 104]. For further details concerning the theory and the applications of long-range dependence from a more statistical point of view, we also refer to [31]. The fBm was first introduced within a Hilbert space framework by Kolmogorov in 1940 in [64], where it was called Wiener Helix. It was further studied by Yaglom in [118]. The name fractional Brownian motion is due to Mandelbrot and Van Ness, who in 1968 provided in [74] a stochastic integral representation of this process in terms of a standard Brownian motion.

Definition. A centered Gaussian process $W = \left(W_t^H \,:\, t \geq 0\right)$, $H \in (0,1)$ is called standard fractional Brownian motion of Hurst parameter H if it has the covariance function

$$R_H(t,s) := \mathbb{E}\left(W_t^H W_s^H\right) = \frac{1}{2}\left\{t^{2H} + s^{2H} - |t-s|^{2H}\right\}, \quad t, s \geq 0 \tag{2.24}$$

For $H = 1/2$, the fBm is then a standard Bm known as the Wiener process. For $H \neq 1/2$ the fBm is niether a semimartingale nor a Markov process. By definition, it is easy to see that fBm has the following properties

1. $\mathbb{P}(W_0^H = 0) = 1$, $\mathbb{E}(W_t^H) = 0$ and $\mathbb{E}[(W_t^H)^2] = t^{2H}$ for all $t \geq 0$.
2. The fBm W^H has homogeneous increments, i.e., $W_{t+s}^H - W_s^H$ has the same law of W_t^H, for all $t, s \geq 0$.
3. The fBm is self-similar with index H, that is, for any $a > 0$, the processes $\left(a^{-H} W_t^H \,:\, t \geq 0\right)$ and $\left(W_{at}^H \,:\, t \geq 0\right)$ have the same law.
4. The fBm W^H admits a continuous modification, which is locally Hölder with exponent $\gamma \in (0, H)$.
5. The sample paths of the fBm W^H are nowhere differentiable in the L^2-sense.

6. For the fBm $\left(W_t^H \ : \ 0 \le t \le T\right)$, with Hurst parameter $H \in (0,1)$, we have ($\mathbb{P}$-a.s)

$$\begin{aligned}\underset{n\to\infty}{\text{l.i.m}} \sum_{j=0}^{2^n-1} \left| W^H\left(\frac{j+1}{2^n}T\right) - W^H\left(\frac{j}{2^n}T\right)\right|^p &= 0\ , \quad \text{if } pH > 1,\\ &= \infty\ , \quad \text{if } pH < 1,\\ &= T\ , \quad \text{if } pH = 1.\end{aligned}$$

In particular

$$\underset{n\to\infty}{\text{l.i.m}} \sum_{j=0}^{2^n-1} \left| W^H\left(\frac{j+1}{2^n}T\right) - W^H\left(\frac{j}{2^n}T\right)\right|^{1/H} = T$$

2.2.2 Wiener Integrals w.r.t Fractional Brownian Motion

It is known that, in order to develop the theory of stochastic integration of a random process with respect to another stochastic process satisfying the usual properties of integrals such as linearity and dominated convergence theorem, it is necessary for the integrator to be a semimartingale. This can be seen from Theorem VIII.80 in [26]. Semimartingales can also be characterized by this property. Since fBm is not a semimartingale, it is not possible to define stochastic integration of a random process with respect to fBm starting with the usual method of limiting arguments based on Riemann-type sums for simple functions as in the case of Itô integrals. However, the special case of a stochastic integration of a deterministic integrand with respect to fBm as the integrator can be developed using the theory of integration w.r.t general Gaussian processes as given in [57] and more recently in [5]. There are other methods of developing stochastic integration of a random process with respect to fBm using the notion of Wick product and applying the techniques of Malliavin calculus. We do not use these approaches throughout this work.

Let $W^H = \left(W_t^H \ : \ t \in \mathbb{R}\right)$ be a standard fBm with Hurst index $H > 1/2$ and suppose $Y = (Y_t \ : \ t \in \mathbb{R})$ is a simple process in the sense that $Y_t = \sum_{j=1}^{k} \alpha_j \chi_{(T_{j-1},T_j]}(t)$, where $-\infty < T_0 < T_1 < \cdots < T_k < \infty$ and α_j are $\mathcal{F}_{T_j}$-measurable r.v's. We define the stochastic integral of the process Y w.r.t W^H as

$$\int_{\mathbb{R}} Y_t dW_t^H = \sum_{j=1}^{k} \alpha_j \left(W_{T_j}^H - W_{T_{j-1}}^H\right). \tag{2.25}$$

If the process Y is of locally bounded variation, then we can define this integral by using the integration by parts formula

$$\int_s^t Y_u dW_u^H = Y_t W_t^H - Y_s W_s^H - \int_s^t W_u^H dY_u, \quad -\infty < s \le t < \infty, \tag{2.26}$$

and the integral on the RHS of (2.26) can be defined using the theory of Lebesgue-Stieltjes integration. Suppose that the process Y is non random. Under suitable conditions on the non random function Y, the integral on the LHS of (2.26) can be defined as an L^2-limit of Riemann sums of the type defined in (2.25) with nonrandom sequence T_j . Gripenberg and Norros [52] give an example of a random process Y illustrating the problem of non continuity in extending the above method of stochastic integration w.r.t fBm for random integrands. An alternate way of defining the stochastic integral of a non random function Y_t w.r.t fBm W^H is by the formula

$$\int_{\mathbb{R}} Y_t dW_t^H := c_H \int_{\mathbb{R}} \left(\int_u^\infty (t-u)^{H-3/2} Y_t dt \right) dW_u, \tag{2.27}$$

where W is a standard Wiener process and c_H is a nonnegative constant to be specified later. The integral defined on the RHS of (2.27) exists provided the function

$$\int_u^\infty (t-u)^{H-3/2} Y_t dt$$

as a function of u is square integrable. A sufficient condition for this to hold is that $Y_t \in L^2(\mathbb{R}) \cap L^1(\mathbb{R})$. This condition is obviously satisfied by the function $Y_s = \chi_{[0,t]}(s)$, which leads to

$$\begin{aligned} W_t^H &= c_H \int_{-\infty}^0 \left((t-u)^{H-1/2} - (-u)^{H-1/2} \right) dW_u + c_H \int_0^t (t-u)^{H-1/2} dW_u \\ &= c_H \int_{-\infty}^t \left[(t-u)^{H-1/2} - ((-u)^+)^{H-1/2} \right] dW_u := c_H Z_t, \end{aligned} \tag{2.28}$$

where $x^+ = \max\{0, x\}$. The equality (2.28) makes sense as an integral representation of the fBm, if it provides the correct covariance function $R_H(t,s)$ given in (2.24). First note that the RHS of (2.28) is well defined. In fact, for $t \geq 0$, we have

$$\begin{aligned} &\int_{-\infty}^t \left[(t-u)^{H-1/2} - ((-u)^+)^{H-1/2} \right]^2 du \\ &= \int_{-\infty}^0 \left[(t-u)^{H-1/2} - (-u)^{H-1/2} \right]^2 du + \int_0^t (t-u)^{2H-1} du \\ &= t^{2H} \left[\int_0^\infty \left((1+v)^{H-1/2} - v^{H-1/2} \right)^2 du + \frac{1}{2H} \right] = \mathbb{E}(Z_t^2). \end{aligned} \tag{2.29}$$

In the last equality we used change of variables ($v = -u/t$). The integral presented in (2.29) is finite, since

$$\left((1+v)^{H-1/2} - v^{H-1/2} \right)^2 \sim \left(H - \frac{1}{2} \right)^2 v^{2H-3}, \quad \text{as } v \to \infty. \tag{2.30}$$

It is then convenient to set $c_H = \left[\int_0^\infty \left((1+v)^{H-1/2} - v^{H-1/2}\right)^2 du + \frac{1}{2H}\right]^{-1/2}$, so that $\mathbb{E}[(W_t^H)^2] = c_H^2 \mathbb{E}(Z_t^2) = t^{2H}$. Similarly, for any $s < t$, we obtain

$$\mathbb{E}\left|Z_t - Z_s\right|^2 = c_H^{-2}\left|t-s\right|^{2H}.$$

Hence,

$$\begin{aligned}
\mathbb{E}\left(c_H Z_t c_H Z_s\right) &= c_H^2 \mathbb{E}(Z_t Z_s) \\
&= \frac{c_H^2}{2}\left\{\mathbb{E}Z_t^2 + \mathbb{E}Z_s^2 - \mathbb{E}\left|Z_t - Z_s\right|^2\right\} \\
&= R_H(t,s).
\end{aligned}$$

It is easy to see that the random variable $\int_{\mathbb{R}} Y_t dW_t^H$ is a Gaussian with zero mean, whenever $Y_t \in L^2(\mathbb{R}) \cap L^1(\mathbb{R})$. The following theorem provides more general covariance formula, for which $R_H(t,s)$ is a special case. The proof of such a formula can be found in [52, 94]. Throughout this thesis we restrict ourselves to the range $H \in (1/2, 1)$ and the stochastic integrals of random functions $\{u_t \ : \ t \in [0,T]\}$ should be understood in path-wise sense, whenever the integrands $(u_t)_t$ have γ-Hölder continuous trajectories with $\gamma > 1 - H$. The existence of such integrals is justified by Young's results [119].

Theorem 2.2.1. *For nonrandom functions $Y_t^1, Y_t^2 \in L^2(\mathbb{R}) \cap L^1(\mathbb{R})$, we have*

$$\mathbb{E}\left(\int_{\mathbb{R}} Y_t^1 dW_t^H \int_{\mathbb{R}} Y_t^2 dW_t^H\right) = H(2H-1)\int\int_{\mathbb{R}^2} Y_t^1 Y_s^2 \left|t-s\right|^{2H-2} dtds.$$

In most of our non-parametric procedures, we used the following result which is due to Mémin *et al.*,[77].

Theorem 2.2.2. *Let $\{Y_t \ : \ t \in [0,T]\}$ be a non random function such that $\int_a^b Y_t dW_t^H$ exists. There exist two nonnegative constants $C(H,r)$ and $c(H,r)$ such that for every $r > 0$ and $0 \leq a < b < \infty$, we have*

$$\mathbb{E}\left|\int_a^b Y_t dW_t^H\right|^r \leq C(H,r)\left\|Y_\cdot\right\|_{L^{1/H}(a,b)}^r, \ \textit{for all } H > 1/2$$

and

$$\mathbb{E}\left|\int_a^b Y_t dW_t^H\right|^r \geq c(H,r)\left\|Y_\cdot\right\|_{L^{1/H}(a,b)}^r, \ \textit{for all } H < 1/2.$$

2.2.3 Stochastic Differential Equations driven by fBm

Consider the functions $A(t,x),\ B(t,x)\ :\ [0,T]\times\mathbb{R}\longrightarrow\mathbb{R}$ satisfying the following assumptions

(i) The function $B(t,x)$ is differentiable in x with derivative $B_x(t,x)$, there exist $M>0$, $0<\gamma,\ \kappa\le 1$ and for any $R>0$, there exists $M_R>0$ such that

$$\begin{aligned}|B(t,x)-B(t,y)| &\le M \quad |x-y| \quad , \quad \text{for all } t\in[0,T],\ x,y\in\mathbb{R}\\ |B_x(t,x)-B_x(t,y)| &\le M_R \quad |x-y|^\kappa \quad , \quad \text{for all } t\in[0,T],\ |x|,|y|\le R\\ |B(t,x)-B(s,x)|+|B_x(t,x)-B_x(s,x)| &\le M\,|t-s|^\gamma, \quad \text{for all } s,t\in[0,T],\ x\in\mathbb{R}.\end{aligned}$$

(ii) For any $R>0$ there exists $L_R>0$ such that

$$|A(t,x)-A(t,y)|\le L_R\,|x-y|\,, \quad \text{for all } t\in[0,T],\ |x|,|y|\le R$$

(iii) There exists a function $A_0(t)\in L^p([0,T])$ and $L>0$ such that $|A(t,x)|\le L\,|x|+A_0(t)$ for all $(t,x)\in[0,T]\times\mathbb{R}$

Theorem 2.2.3. ***(Mishura [80, p. 201])*** *Let the coefficients $A(\cdot)$, $B(\cdot)$ satisfy the assumptions **(i)-(iii)** with $p=(1-H+\varepsilon)^{-1}$, $0<\varepsilon<H-1/2$, $\gamma>1-H$, $\kappa>H^{-1}-1$ (the constants M, M_R, R, L_R and the funcion $A_0(t)$ can depend on ω). Then the following SDE*

$$X_t = X_0+\int_0^t A(s,X_s)ds+\int_0^t B(s,X_s)dW_s^H,\ t\in[0,T],\ H\in(1/2,1). \tag{2.31}$$

admits a unique strong solution $X=(X_t\ :\ 0\le t\le T)$ with trajectories from $C^{H-\varepsilon}([0,T])$ $\mathbb{P}$-a.s.

Remark. Theorem 2.2.3 admits evident generalization to multidimensional case. Consider the SDE on $\mathbb{R}^d$ given by

$$X_t^i = X_0^i+\int_0^t A_i(s,X_s)ds+\sum_{j=1}^m\int_0^t B_{i,j}(s,X_s)dW_s^{H_j},\ 1\le i\le d,\ 0\le t\le T. \tag{2.32}$$

where W^{H_j} are fBm's with Hurst parameters $H_j\in(1/2,1)$, $j=1,\cdots,m$. Denote by $B=(B_{i,j})_{i,j=1}^{d,m}$ the matrix of "diffusions"and $A=(A_i)_{i=1}^d$ the "drift "vector, $|A|:=\left(\sum_{i=1}^d A_i^2\right)^{1/2}$, $|B|:=\left(\sum_{i,j}B_{i,j}^2\right)^{1/2}$, and suppose that assumptions **(i)-(iii)** hold with these notations, $H=\min_{1\le j\le m}H_j$. Then there exists a unique vector solution X_t of equation (2.32) with trajectories from $C^{H-\varepsilon}([0,T])$ $\mathbb{P}$-a.s.

CHAPTER 3
LITERATURE REVIEW AND THE MODEL

Parameteric and nonparameteric estimation in the context of random effects models has ben recently investigated by many authors (see, e.g. [29, 25, 89, 90, 83, 82, 81, 6, 22]). In these models, the noise is represented by a Brownian motion known by independence property of its increments. Such a property may not be valid for many phenomena arising in a variety of different scientific fields, including hydrology [75], biology [21], medicine [68], economics [50] or traffic network [117]. As a result self-similar processes have been used to successfully model data exhibiting long-range dependence. Among the simplest models that display long-range dependence, one can consider the fractional Brownian motion, introduced in the statistics community by Mandelbrot and Van Ness[74]. A normalized fBm with the Hurst index $H \in (0,1)$ is centered Gaussian process $\left(W_t^H,\ t \geq 0\right)$ having the covariance defined by (2.24).

In modeling, the problems of statistical estimation of model parameters are of particular importance, so the growing number of papers devoted to statistical methods for equations with fractional noise is not surprising. We will cite only few of them; further references can be found in [80] and [94]. In [63], the authors proposed and studied maximum likelihood estimators for fractional OrnsteinUhlenbeck process. Related results were obtained in [95], where a more general model was considered. In [56] the authors proposed a least squares estimator for fractional OrnsteinUhlenbeck process and proved its asymptotic normality.

It is worth to mention that papers [55] and [112] deal with the whole range of Hurst parameter $H \in (0,1)$, while other papers cited here investigate only the case $H > 1/2$ (which corresponds to long range dependence); recall that in the case $H = 1/2$, we have a classical diffusion, and there is a huge literature devoted to it; we refer to books: [69], and [73](Vol II) for the review of the topic. Chepter 4 is concerned with this case, for which we consider a nonlinear homogeneous model with generalized random effects. In this chapter both consistensy and asymptotic normality of our proposed estimators are established and illustrated by various examples.

In the context of stochastic differential equation models with random effects, which are increasingly used in the biomedical field and have proved to be adequate tools for the study of repeated measurements collected on a series of subjects, parametric and non parametric inference has recently been investigated by many authors (see, e.g.[6, 22, 89, 90]). However, there is no reference at present related to inference for SDE's with random effects driven by fBm, except our recent papers [36, 37, 38]. Chapter 5 is based on these papers, and concerns different models driven by fBm and treated by different approachs. Compared to the parametric framework, the problem of non parametric estimation for FSDE's has gained less attention. For the theoretical study of this problem, the main contribution to our knowledge is due to Prakasa[93]. In[79], the authors studied the problem of nonparametric estimation for a fractional process with small diffusion, when $H \in (1/2, 3/4)$ as $\varepsilon \to 0$. An other recent work[92] tackled the same problem when the fractional Brownian motion is replaced by a mixed one.

In this thesis we are concerned with a general model described by N stochastic processes of the following form:

$$\begin{aligned} dX^i(t) &= \Sigma(t, X^i(t), \phi_i)dt + \Lambda(t, X^i(t))dW^{H,i}(t) \\ X^i(0) &= x^i \in \mathbb{R},\ i = 1, \cdots, N, \end{aligned}$$

where ϕ_i are random effects with common density (parametric or not) to be estimated. $\Sigma(\cdot)$ and $\Lambda(\cdot)$ denote here the drift and the diffusion terms. In each chapter (Chapter 4 & 5) the drift and the diffusion are specified, and existence and uniqueness of the processes X^i are discussed. For that purpose, Theorems 2.1.10-2.1.11 and Theorem 2.2.3 are used in Chapter 4 & 5 respectively. The following theorems are of great interest and systematically used in Chapter 4 to establish the asymptotic behaviour of the proposed estimators.

Theorem 3.0.4. ***(Schervish (1995)[101, p. 415])*** *Let $\{X_n\}_{n=1}^{\infty}$ be conditionally i.i.d given θ with density $f_{X_1|\Theta}(.|\theta)$ w.r.t a measure ν on a space $(\mathcal{X}^1, \mathcal{B}^1)$. Fix $\theta_0 \in \Theta$ and define, for each $M \subseteq \Theta$ and $x \in \mathcal{X}^1$,*

$$Z(M, x) = \inf_{\psi \in M} \log \frac{f_{X_1|\Theta}(x|\theta_0)}{f_{X_1|\Theta}(x|\psi)}$$

Assume that for each $\theta \neq \theta_0$ there is an open set N_θ such that $\theta \in N_\theta$ and $\mathbb{E}_{\theta_0} Z(N_\theta, X_i) > 0$. If Θ is not compact, assume further that there is a compact $C \subseteq \Theta$ such that $\theta_0 \in C$ and $\mathbb{E}_{\theta_0} Z(\Theta \setminus C, X_i) > 0$. Then, $\lim \widehat{\theta}_n = \theta_0$, a.s. $[\mathbb{P}_{\theta_0}]$.

Theorem 3.0.5. ***(Schervish (1995)[101, p. 418])*** *Assume the same conditions as in Theorem 3.0.4, Except that we now only require that $\mathbb{E}_{\theta_0} Z(N_\theta, X_i) > -\infty$. Assume further that $f_{X_1|\Theta}(x|\theta)$ is continuous in θ for every x, a.s $[\mathbb{P}_{\theta_0}]$. Then, $\lim \widehat{\theta}_n = \theta_0$, a.s. $[\mathbb{P}_{\theta_0}]$.*

Theorem 3.0.6. ***(Schervish (1995)[101, p. 421])*** *Let Θ be a subset of $\mathbb{R}^p$, and let $\{X_n\}_{n=1}^{\infty}$ be conditionally i.i.d given θ with density $f_{X_1|\Theta}(.|\theta)$. Let $\hat{\theta}_n$ be an MLE. Assume* *that* $\widehat{\theta}_n \xrightarrow{P} \theta$ *under* $\mathbb{P}_\theta$ *for all* θ. *Assume that* $f_{X_1|\Theta}(x|\theta)$ *has continuous second partial derivatives with respect to* θ *and that differentiation can be passed under the integral sign. Assume that there exists* $H_r(x,\theta)$ *such that, for each* $\theta_0 \in \dot{\Theta}$ *(in the interior) and for each* $k,\ j$,

$$\sup_{\|\theta-\theta_0\|\le r}\left|\frac{\partial^2}{\partial_{\theta_k}\partial_{\theta_j}}\log f_{X_1|\Theta}(x|\theta_0) - \frac{\partial^2}{\partial_{\theta_k}\partial_{\theta_j}}\log f_{X_1|\Theta}(x|\theta)\right| \le H_r(x,\theta_0),$$

with

$$\lim_{r\to 0}\mathbb{E}_{\theta_0} H_r(X,\theta_0) = 0.$$

Assume that the Fisher Information matrix $\mathcal{I}_{X_1}$ *is finite and nonsingular. Then, under* $\mathbb{P}_{\theta_0}$,

$$\sqrt{N}\left(\widehat{\theta}_N - \theta_0\right) \overset{\mathcal{L}}{\Longrightarrow} \mathcal{N}\left(0, \mathcal{I}_{X_1}^{-1}(\theta_0)\right).$$

CHAPTER 4
STATISTICAL INFERENCE FOR STOCHASTIC PROCESSES WITH GENERALIZED RANDOM EFFECTS

The goal of this chapter is to provide statistical methods of parametric estimation for SDE's given by

$$\begin{aligned} dX^i(t) &= H_1\left(X^i(t),\phi_i\right)dt + H_2\left(X^i(t)\right)dW^i(t), \quad t \le T, \\ X^i(0) &= x_i \in \mathbb{R}, \quad i = 1,\cdots,N, \end{aligned} \tag{4.1}$$

where ϕ_i are random variables (random effects) with common density $g(\varphi,\theta)$ and $\theta \in \Theta$ (compact set) are parameters to be estimated. The results we provide here extend those previously available in the literature. This chapter is split into two sections. First, we derive a class of estimators $\widehat{\theta}_N^{(\varepsilon)}$ converging toward the true value θ_0, as $N \to \infty$ for a nonlinear drift coefficient with generalized random effects. Asymptotic properties of the proposed estimators are established and illustrated by numerical examples as well. Second, we try to weaken the assumptions and show the asymptotic results for the case $\varepsilon = 0$. To do this, we have expanded the likelihood function by means of iterated Itô integrals.

4.1 PARAMETRIC ESTIMATION OF THE POPULATION PARAMETERS

4.1.1 Problem Outline

Consider N subjects $X^1,\cdots,X^N$ defined on a complete probability space $(\Omega,\mathcal{F},\mathbb{P})$ by (4.1). Let $\left(\mathcal{F}_t^i\right)_{t\ge 0}$ be a collection of nondecreasing families of sub-σ-algebras of $\mathcal{F}$ and $\left(W^i,\mathcal{F}_t^i,\ t\le T\right)$, $i = 1,\cdots,N$ are independent Wiener processes. Let also $\phi_1,\cdots,\phi_N$ be N i.i.d $\mathbb{R}^p$-valued random variables on the common probability space $(\Omega,\mathcal{F},\mathbb{P})$ independent of $\left(W^i,\mathcal{F}_t^i,\ t\le T\right)$. We introduce assumptions ensuring that the processes (4.1) are well defined and allow us to compute the likelihood function based on the observations. Consider the family of sub-σ-algebras of $\mathcal{F}$, defined by $\sigma\left(\phi_j, j\le N\right)\bigvee \mathcal{F}_t^i$. Each W^i is $\mathcal{F}_t$-Brownian motion. Furthermore, the random variables ϕ_i are $\mathcal{F}_0$-measurable. Denote by $g(\varphi,\theta)d\nu(\varphi)$ the common density of the variables $\phi_1,\cdots,\phi_N$, where ν is some dominating measure on $\mathbb{R}^p$, and θ is unknown parameter belonging to a set $\Theta\subset\mathbb{R}^d$, with interior $\dot{\Theta}$ containing the true parameter θ_0. We also assume that Θ is compact and the function $H_2(x)$ has a support such that, $\mathbb{R}\setminus \mathrm{supp}H_2$ is countable set in $\mathbb{R}$.

$\mathbf{A}_1$: The functions $H_1(x,\varphi)$ and $H_2(x)$ are nonanticipative functionals satisfying the local lipschitz condition: For any $R<\infty$ and $\max\{|x|+\|\varphi\|\,;|\widetilde{x}|+\|\widetilde{\varphi}\|\}\le R$ there exists a constant $L_R>0$ such that

$$|H_1(x,\varphi)-H_1(\widetilde{x},\widetilde{\varphi})|^2+|H_2(x)-H_2(\widetilde{x})|^2\le L_R\left(|x-\widetilde{x}|^2+\|\varphi-\widetilde{\varphi}\|^2\right),$$

and satisfying also the following monotone condition: There exists a nonnegative constant K such that for all $(x,\varphi)\in\mathbb{R}\times\mathbb{R}^p$

$$xH_1(x,\varphi)+\frac{1}{2}|H_2(x)|^2\le K\left(1+|x|^2+\|\varphi\|^2\right).$$

Under $\mathbf{A}_1$, the system (4.1) admits a unique continuous strong solution $\big((X^i,\phi^i);\ t\ge 0\big)$, with probability one (see, Theorem 2.1.11). Moreover, there exists a function $h_i\ :\ \mathbb{R}\times\mathbb{R}^p\times C(\mathbb{R}^+,\mathbb{R})$ with $X^i=h_i\left(x^i,\phi_i,W^i\right)$ (see, e.g., [60, p. 310]). Let C_{T_i} denote the space of real continuous functions $(x(t)\ :\ t\in[0,T_i])$ defined on $[0,T_i]$ endowed with σ-field $\mathcal{B}_{T_i}$. The last σ-field is associated with the topology of uniform convergence on $[0,T_i]$. Under $\mathbf{A}_1$, we introduce the distribution $\mu_{X^i_{\varphi,x^i}}$ on $(C_{T_i},\mathcal{B}_{T_i})$ of the process $\big(X^i,x^i|\phi_i=\varphi\big)$. On $\mathbb{R}^p\times C_{T_i}$, $Q^i_\theta=g(\varphi,\theta)d\nu(\varphi)\otimes\mu_{X^i_{\varphi,x^i}}$ denotes the joint distribution of (ϕ_i,X^i) and $\mathbb{P}^i_\theta$ denotes the marginal distribution of $\big(X^i(t);0\le t\le T_i\big)$ on $(C_{T_i},\mathcal{B}_{T_i})$. Since $\mathbb{R}\setminus\operatorname{supp}H_2$ is countable, for all $\varphi\neq\widetilde{\varphi}$, we may define the process $\alpha^i_t(\omega)$, $t\le T_i$ by $\alpha^i_t=H_2^+(X^i(t))\left(H_1(X^i(t),\varphi)-H_1(X^i(t),\widetilde{\varphi})\right)$, where

$$h^+(x)=\begin{cases}\dfrac{1}{h(x)}\,, & h(x)\neq 0,\\[2ex] 0\,, & h(x)=0.\end{cases}$$

Let us consider the following assumptions insuring the equivalence of the measures $(\mu_{X^i_{\varphi,x^i}},\ \varphi\in\mathbb{R}^p)$ and $\mu_{X^i_{\varphi_0,x^i}}$, where φ_0 is fixed.

$\mathbf{A}_2$: $\mathbb{E}\left(\mathbf{e}^{-\int_0^{T_i}\alpha^i_t dW^i(t)-\frac{1}{2}\int_0^{T_i}{\alpha^i_t}^2dt}\right)=1,\quad i=1\cdots,N.$

$\mathbf{A}_3$: For $i=1\cdots,N$ and for all $\varphi\in\mathbb{R}^p$,

$$\mathbb{P}\left(\int_0^{T_i}\left[H_2^+(X^i(t))H_1(X^i(t),\varphi)\right]^2dt<\infty\right)=1\ .$$

Since the individuals X^i are independent (this is inherited from the independence of ϕ_i and W^i). The distribution of the whole sample $(X^i(t),t\le T_i;i=1,\cdots,N)$ on $C=\prod_{i=1}^{N}C_{T_i}$ is defined by $\mathbb{P}_\theta = \otimes_{i=1}^N \mathbb{P}_\theta^i$. Now we can define the likelihood function as

$$\Lambda(\theta) = \frac{d\mathbb{P}_\theta}{d\mathbb{P}} = \prod_{i=1}^N \frac{d\mathbb{P}_\theta^i}{d\mathbb{P}^i} \text{ , where } \mathbb{P} = \otimes_{i=1}^N \mathbb{P}^i \text{ and } \mathbb{P}^i = \mu_{X^i_{\varphi_0, x^i}}.$$

4.1.2 Contrast Functions

In this subsection, we construct the likelihood function from which we derive a class of contrast functions. We have the following propositions for which the proof is relegated to the Appendix A.

Proposition 4.1.1. *Let the conditions* $\boldsymbol{A}_1$- $\boldsymbol{A}_3$ *be satisfied and let* $\varphi, \varphi_0 \in \mathbb{R}^p$. *Then* $\mu_{X^i_{\varphi,x^i}} \sim \mu_{X^i_{\varphi_0,x^i}}$, *and* $\dfrac{d\mu_{X^i_{\varphi,x^i}}}{d\mu_{X^i_{\varphi_0,x^i}}}(X^i) = L_{T_i}(X^i, \varphi, \varphi_0) = e^{l_{T_i}(X^i, \varphi, \varphi_0)}$, *where*

$$\begin{aligned} l_{T_i}(X^i, \varphi, \varphi_0) &= \int_0^{T_i} H_2^+(X^i(s))^2 \left[H_1(X^i(s), \varphi) - H_1(X^i(s), \varphi_0)\right] dX^i(s) \\ &\quad - \frac{1}{2}\int_0^{T_i} H_2^+(X^i(s))^2 \left[H_1^2(X^i(s), \varphi) - H_1^2(X^i(s), \varphi_0)\right] ds. \end{aligned}$$

Moreover, the exact likelihood of the whole sample $(X^i(t); t \in [0, T_i],\ i = 1 \cdots, N)$ *can be expressed as*

$$\Lambda(\theta) = \prod_{i=1}^N \Lambda_i(X^i, \theta),$$

where

$$\Lambda_i(X^i, \theta) = \frac{d\mathbb{P}_\theta^i}{d\mathbb{P}^i} = \int L_{T_i}(X^i, \varphi, \varphi_0) g(\varphi, \theta) d\nu(\varphi).$$

Denote by $\Gamma(\theta)$ the log-likelihood of the whole sample $(X^i(t); t \in [0, T_i],\ i = 1 \cdots, N)$.

Proposition 4.1.2. *For* $i = 1, \cdots, N$ *the likelihood function* $\Lambda_i(X^i, \theta)$ *can be expressed as follows:*

$$\Lambda_i(X^i, \theta) = 1 + \int_0^{T_i} \gamma_s(X^i, \theta) dW^i(s)$$

where $X^i \longrightarrow \gamma_s(X^i, \theta)$ *is a measurable function on* C_{T_i} *with explicit expression given in Appendix A (see the proof). Furthermore, if* $\mathbb{E}\displaystyle\int_0^{T_i} \gamma_s^2(X^i, \theta) ds < \infty$. *Then* $\Gamma(\theta)$ *has the following form :*

$$\Gamma(\theta) = \sum_{i=1}^N \int_0^{T_i} \Psi^i(s, \theta) dW^i(s) - \frac{1}{2}\int_0^{T_i} \Psi^i(s, \theta)^2 ds,$$

where $\Psi^i(s,\theta)$ is a function of $\gamma_t(X^i,\theta)$ such that $\mathbb{P}\left(\int_0^{T_i} \Psi^i(t,\theta)^2 dt < \infty\right) = 1$. *An explicit expression of $\Psi^i(s,\theta)$ is given in Appendix A (see the proof).*

For each $\varepsilon > 0$, we consider the contrast function $\Lambda^{(\varepsilon)}(\theta) = \prod_{i=1}^{N} \Lambda_i^{(\varepsilon)}(X^i,\theta)$, where $\Lambda_i^{(\varepsilon)}(X^i,\theta) = 1 + (1+\varepsilon)^{-1} \int_0^{T_i} \gamma_s(X^i,\theta) dW^i(s)$. Our proposed estimators $\widehat{\theta}_N^{(\varepsilon)}$ will be defined as $\widehat{\theta}_N^{(\varepsilon)} = \arg\max_{\theta\in\Theta} \Lambda^{(\varepsilon)}(\theta)$. It is easy to see $\Lambda_i^{(\varepsilon)}(X^i,\theta) \to \Lambda_i(X^i,\theta)$ as $\varepsilon \to 0$. Under some regularity conditions, one can show that $\widehat{\theta}_N^{(\varepsilon)}$ converges to the maximum likelihood estimator as $\varepsilon \to 0$.

4.1.3 Asymptotic Behaviour of $\widehat{\theta}_N^{(\varepsilon)}$

To prove the consistency result of our proposed estimators, we consider for any X^i, $i = 1,\ldots,N$, the random functions $Z_\varepsilon(\widetilde{\Theta}, X^i)$ defined by

$$Z_\varepsilon\left(\widetilde{\Theta}, X^i\right) := \inf_{\theta\in\widetilde{\Theta}} \log \frac{\Lambda_i^{(\varepsilon)}(X^i,\theta_0)}{\Lambda_i^{(\varepsilon)}(X^i,\theta)},$$

where $\widetilde{\Theta} \subset \Theta$. We recall also the identifiability assumption for the marginal densities which is a natural and even a necessary condition i.e., $\Lambda_i(X^i,\theta) = \Lambda_i(X^i,\theta') \Longrightarrow \theta = \theta'$. Let $T_i = T$, $x^i = x$ for $i = 1,\cdots,N$, so that the observed processes $\left(X^i(t), t \in [0,T]\right)$, $i = 1,\cdots,N$, are conditionally i.i.d given θ with common density $\Lambda_1(x,\theta)$ on C_T. In the sequel, we will focus on statistical results under $\mathbb{P}^1_{\theta_0}$. We simplify our notations by setting $\Lambda^{(\varepsilon)}(x,\theta) = \Lambda_1^{(\varepsilon)}(x,\theta)$ and $\mathbb{P}_{\theta_0} = \mathbb{P}^1_{\theta_0}$. Now, with these arguments, the following statements hold :

Theorem 4.1.3. *For any fixed $\varepsilon > 0$, we have*

- **i-** *If for any given $\theta \neq \theta_0$, there is an open set N_θ such that $\theta \in N_\theta$ and $\mathbb{E}_{\theta_0} Z_\varepsilon(N_\theta, X^i) > 0$. Then $\lim \widehat{\theta}_N^{(\varepsilon)} = \theta_0$, almost surely under $\mathbb{P}_{\theta_0}$.*

- **ii-** *Let the assumptions $\boldsymbol{A}_1$- $\boldsymbol{A}_3$ be fulfilled. Assume also that $\Lambda_i^{(\varepsilon)}(x,\theta)$ is continuous in θ $\mathbb{P}_{\theta_0}-a.s.$ and*

$$\mathbb{E}_{\theta_0} \int_0^{T_i} \gamma_s^2(X^i,\theta) ds < \infty, \ \text{for } i = 1,\cdots,N \text{ all } \theta. \tag{4.2}$$

For each $\theta \neq \theta_0$ there is an open set N_θ such that $\theta \in N_\theta$ and $\mathbb{E}_{\theta_0} Z_\varepsilon(N_\theta, X^i) > -\infty$.

Proof. To prove this theorem, we follow Schervich (see, [101, Section 7.3]). So for the first statement, we shall prove that for every $\delta > 0$,

$$\mathbb{P}_{\theta_0}\left(\limsup_{N\to\infty}\left\|\widehat{\theta}_N^{(\varepsilon)} - \theta_0\right\| \geq \delta\right) = 0. \tag{4.3}$$

Let $\delta > 0$, and let N_{θ_0} be an open ball of radius ε centered at θ_0. The set $\Theta \setminus N_{\theta_0}$ is compact with cover family $\left\{N_\theta : \theta \in \Theta \setminus N_{\theta_0},\ \mathbb{E}_{\theta_0}(Z_\varepsilon(N_\theta, X^i)) > 0\right\}$ from which we extract the following finite subcover $N_{\theta_1}, \ldots, N_{\theta_m}$ such that $\Theta = \cup_{j=0}^m N_{\theta_j}$.

Let $C_T = \chi$ denote the space of real continuous functions $(x(s) : s \in [0,T])$ defined on $[0,T]$, endowed with σ-field $\mathcal{C}_T$ associated with the topology of uniform convergence on $[0,T]$. Let χ^∞ be the infinite product space of copies of χ. Let $x = (x_1, x_2, \ldots) \in \chi^\infty$ denotes a sequence of possible values of $(X^1, X^2, \ldots)$. By the strong law of large numbers, $N^{-1}\sum_{i=1}^{N} Z_\varepsilon(N_{\theta_j}, X^i)$ converges to $\mathbb{E}_{\theta_0}(Z_\varepsilon(N_{\theta_j}, X^i)) = c_j > 0$ $\mathbb{P}_{\theta_0}$-a.s, for all $j = 1, \ldots, m$. Consider $B_j \subset \chi^\infty$ the set of data sequences such that convergence holds, and set $B = \cap_{j=1}^m B_j$. Then $\mathbb{P}_{\theta_0}(B) = 1$. For each $x \in B$ we have also :

$$\begin{aligned}
\left\{x : \limsup_{N\to\infty}\left\|\widehat{\theta}_N^{(\varepsilon)}(x) - \theta_0\right\| \geq \delta\right\} &\subset \bigcup_{j=1}^m \left\{x\ :\ \widehat{\theta}_N^{(\varepsilon)}(x) \in N_{\theta_j},\ \text{ infinitely often}\right\} \\
&\subset \bigcup_{j=1}^m \left\{x : \frac{1}{N}\sum_{i=1}^N Z_\varepsilon(N_{\theta_j}, x_i) \leq 0,\ \text{infinitely often}\right\} \\
&\subset \bigcup_{j=1}^m B_j^c = B^c.
\end{aligned}$$

where A^c denotes the complement of A in χ^∞. Since $\mathbb{P}_{\theta_0}(B^c) = 0$, (4.3) follows and the proof of **i-** is complete. **ii-** Let $\theta \neq \theta_0$ and $U_\theta \subset \widetilde{\Theta} = \Theta \setminus \{\theta_0\}$ be an open set, which we specify later. We have

$$\begin{aligned}
\mathbb{E}_{\theta_0} Z_\varepsilon(U_\theta, X^i) &= \mathbb{E}_{\theta_0}\log\Lambda^{(\varepsilon)}(X^i, \theta_0) + \mathbb{E}_{\theta_0} z_{i,\varepsilon}(U_\theta) \\
&\geq \log\left(\frac{\varepsilon}{1+\varepsilon}\right) + \mathbb{E}_{\theta_0} z_{i,\varepsilon}(U_\theta),
\end{aligned}$$

where $z_{i,\varepsilon}(U) = \inf_{\theta'\in U}\left[-\log\Lambda^{(\varepsilon)}(X^i, \theta')\right]$. In the last inequality, we used the fact that $\Lambda^{(\varepsilon)}(X^i, \theta_0) = \dfrac{\varepsilon + \Lambda(X^i, \theta_0)}{1+\varepsilon} \geq \dfrac{\varepsilon}{1+\varepsilon}$. Let $k \geq k_0 > 0$ such that $N_\theta^{(k)} = \overline{B(\theta, 1/k)} \subseteq \widetilde{\Theta}$, where $B(x, \delta)$ denotes the open ball centred at x with radius δ. Since $\Lambda^{(\varepsilon)}(X^i, \theta)$ is continuous in θ on the compact set $N_\theta^{(k)}$, there exists $\theta_k \in N_\theta^{(k)}$ such that $z_{i,\varepsilon}(N_\theta^{(k)}) = -\log\Lambda^{(\varepsilon)}(X^i, \theta_k)$. By the continuity of $\Lambda^{(\varepsilon)}(X^i, \theta)$ and the fact that $\theta_k \longrightarrow \theta$, we have

$$z_{i,\varepsilon}(N_\theta^{(k)}) \longrightarrow -\log\Lambda^{(\varepsilon)}(X^i, \theta) \text{ as } k \longrightarrow \infty. \tag{4.4}$$

Since $N_\theta^{(k+1)} \subseteq N_\theta^{(k)} \subseteq \widetilde{\Theta}$, it follows that $z_{i,\varepsilon}(N_\theta^{(k)}) \geq z_{i,\varepsilon}(\widetilde{\Theta})$. Applying Fatou lemma to the sequence $\left\{z_{i,\varepsilon}(N_\theta^{(k)}) - z_{i,\varepsilon}(\widetilde{\Theta})\right\}_{k=k_0}^{\infty}$ and (4.4), we get

$$\liminf_{k\to\infty} \mathbb{E}_{\theta_0} z_{i,\varepsilon}(N_\theta^{(k)}) \geq \mathbb{E}_{\theta_0} \liminf_{k\to\infty} z_{i,\varepsilon}(N_\theta^{(k)}) = -\mathbb{E}_{\theta_0} \log \Lambda^{(\varepsilon)}(X^i, \theta). \tag{4.5}$$

The condition (4.2) guarantees the martingale property of $\Lambda^{(\varepsilon)}(X^i,\theta)$ with $\mathbb{E}_{\theta_0}\Lambda^{(\varepsilon)}(X^i,\theta) = 1$. Hence, by using Jensen inequality, we obtain

$$-\mathbb{E}_{\theta_0} \log \Lambda^{(\varepsilon)}(X^i,\theta) \geq -\log \mathbb{E}_{\theta_0}\Lambda^{(\varepsilon)}(X^i,\theta) = 0 \tag{4.6}$$

From (4.5) and (4.6), we can choose k^* such that $\mathbb{E}_{\theta_0} z_{i,\varepsilon}(N_\theta^{(k^*)}) \geq 0$. Then, we set $U_\theta = B(\theta, 1/(k^*+1))$ to complete the proof. $\square$

Let us introduce an alternative of the Kullback-Leibler information defined by $\mathcal{J}_X^{(\varepsilon)}(\theta_0,\theta) = \mathbb{E}_{\theta_0} \log \dfrac{\Lambda^{(\varepsilon)}(X^1,\theta_0)}{\Lambda^{(\varepsilon)}(X^1,\theta)}$. We have (see Appendix A. (p.104) for the proof),

$$\mathcal{J}_X^{(\varepsilon)}(\theta_0,\theta) > 0, \quad \text{for all } \theta \neq \theta_0. \tag{4.7}$$

Theorem 4.1.4. *Let the assumptions of* ***ii-*** *in Theorem 4.1.3 be satisfied. Then for each $\theta \neq \theta_0$ there is an open set N_θ containing θ such that $\mathbb{E}_{\theta_0} Z_\varepsilon(N_\theta, X^i) > 0$. In particular the estimator $\widehat{\theta}_N^{(\varepsilon)}$ is strongly consistent.*

Proof. For each $\theta \neq \theta_0$ in Θ, let $N_\theta^{(k)}$ be a closed ball centred at θ with radius $\dfrac{1}{k}$ such that, for each k, $N_\theta^{(k+1)} \subseteq N_\theta^{(k)} \subseteq \widetilde{\Theta}$, where $\widetilde{\Theta}$ is an open set such that $\theta \in \widetilde{\Theta}$ and $\mathbb{E}_{\theta_0} Z_\varepsilon(\widetilde{\Theta}, X^i) > -\infty$. Clearly, $\displaystyle\bigcap_{k=1}^{\infty} N_\theta^{(k)} = \{\theta\}$ and for each X^i, $Z_\varepsilon(N_\theta^{(k)}, X^i)$ increases with k. For each X^i, the function $\Lambda^{(\varepsilon)}(X^i,\theta)$ is continuous on the compact set $N_\theta^{(k)}$. Thus, for each k, we choose $\theta_k \in N_\theta^{(k)}$ so that $Z_\varepsilon\left(N_\theta^{(k)}, X^i\right) = \log \dfrac{\Lambda^{(\varepsilon)}(X^i,\theta_0)}{\Lambda^{(\varepsilon)}(X^i,\theta_k)}$ (the sequence $\{\theta_k\}_{k=1}^{\infty}$ may depend on X^i and ε). Since $\theta_k \longrightarrow \theta$, then

$$\lim_{k\to\infty} Z_\varepsilon\left(N_\theta^{(k)}, X^i\right) = \log \frac{\Lambda^{(\varepsilon)}(X^i,\theta_0)}{\Lambda^{(\varepsilon)}(X^i,\theta)}. \tag{4.8}$$

From the fact that $N_\theta^{(k+1)} \subseteq N_\theta^{(k)}$, we deduce that $Z_\varepsilon(N_\theta^{(k)}, X^i) \geq Z_\varepsilon(\widetilde{\Theta}, X^i)$. If $\mathbb{E}_{\theta_0} Z_\varepsilon(\widetilde{\Theta}, X^i) = \infty$, then our desired result holds with $N_\theta = \widetilde{\Theta}$. If $\mathbb{E}_{\theta_0} Z_\varepsilon(\widetilde{\Theta}, X^i) < \infty$, then we apply Fatou lemma to the sequence $\{Z_\varepsilon(N_\theta^{(k)}, X^i) - Z_\varepsilon(\widetilde{\Theta}, X^i)\}_{k=1}^{\infty}$ and use (4.8) to get

$$\liminf_{k\to\infty} \mathbb{E}_{\theta_0} Z_\varepsilon(N_\theta^{(k)}, X^i) \geq \mathbb{E}_{\theta_0} \liminf_{k\to\infty} \left(Z_\varepsilon(\widetilde{\Theta}, X^i)\right) = \mathcal{J}_X(\theta_0,\theta) > 0.$$

Now, we can choose k^* so that $\mathbb{E}_{\theta_0} Z_\varepsilon\left(N_\theta^{k^*}, X^i\right) > 0$, and apply **i-** in Theorem 4.1.3 to deduce the consistency result. $\square$

Remark. Note that the condition (4.2) is only used to show that $\mathbb{E}_{\theta_0} Z_\varepsilon(N_\theta, X^i) > -\infty$. This condition may be restrictive in some particular cases, where the contrast functions are explicit (see Examples: 4.1.2, 4.1.4 and 4.1.5). However, it is of great interest for general cases, since it allows us to avoid the complexity of integration. (4.2) can be replaced by the following condition and the consistency still holds true:

$$\mathbb{E}_{\theta_0} \sup_{U_\theta} \log \Lambda^{(\varepsilon)}(X^i, \theta) < \infty, \text{ where } U_\theta \text{ is an open set containing } \theta.$$

Example 4.1.1. Consider the following SDE's given by

$$dX^i(t) = \phi_i b(X^i(t))dt + \sigma(X^i(t))dW^i(t); \quad X^i(0) = x \quad, i = 1, \cdots, N,$$

where ϕ_i are random variables with common density $g(\varphi, \theta)d\nu(\varphi) = \mathcal{N}\left(\mu, \omega^2\right)$, $\theta = (\mu, \omega^2) \in \Theta$, and $\Theta = [\underline{\mu}, \overline{\mu}] \times [\underline{\omega^2}, \overline{\omega^2}] \subset \mathbb{R} \times (0, \infty)$. Let $\beta(x) = b(x)/\sigma(x)$ such that $|\beta(x)| \leq C(1 + |x|^{2\gamma})$ for some nonnegative constants γ, C. Assume also that $\mathbb{E}\,|\phi_i|^{2k} < \infty$, for all $k \geq 1$. Simple computations lead to

$$\begin{aligned}\Lambda(X^i, \theta) &= \left[\omega^2 \left(\int_0^T \beta(X^i(t))^2 dt + 1/\omega^2\right)\right]^{-1/2} \\ &\times \exp\left[-\mu^2/2\omega^2 + \left(\int_0^T \beta(X^i(t))dW^i(t) + \mu/\omega^2\right)^2 /2 \left(\int_0^T \beta(X^i(t))^2 dt + 1/\omega^2\right)\right]\end{aligned}$$

Clearly, $\Lambda^{(\varepsilon)}(X^i, \theta)$ is continuous in θ, $\mathbb{P}_{\theta_0}$-a.s. Let $U_\theta = \Theta \backslash \{\theta_0\}$ and $\theta' = (\mu', \omega'^2)$. Using the fact that for every $\varepsilon > 0$, there exists a constant C_ε such that $\log\left(\dfrac{\varepsilon + x}{\varepsilon + 1}\right) \leq C_\varepsilon + (\log(x))^2$, for all $x > 0$, we obtain

$$\begin{aligned}
\mathbb{E}_{\theta_0} \sup_{\theta' \in U_\theta} \log \Lambda^{(\varepsilon)}(X^i, \theta') &= \mathbb{E}_{\theta_0} \sup_{\theta' \in U_\theta} \log \left(\frac{\varepsilon + \Lambda(X^i, \theta')}{\varepsilon + 1} \right) \\
&\leq C_\varepsilon + \mathbb{E}_{\theta_0} \sup_{\theta' \in U_\theta} \left(\log \Lambda(X^i, \theta') \right)^2 \\
&\leq C_\varepsilon + \frac{1}{4} \mathbb{E}_{\theta_0} \sup_{\theta' \in U_\theta} \left\{ - \log \left(\omega'^2 \int_0^T \beta(X^i(t))^2 dt + 1 \right) - \frac{\mu'^2}{\omega'^2} \right. \\
&\qquad \left. + \frac{\left(\int_0^T \beta(X^i(t)) dW^i(t) + \mu'/\omega'^2 \right)^2}{\int_0^T \beta(X^i(t))^2 dt + 1/\omega'^2} \right\}^2 \\
&\leq C_\varepsilon + \frac{3}{4} \left\{ \frac{\overline{\mu}^4}{(\underline{\omega^2})^2} + (\overline{\omega^2})^2 \mathbb{E}_{\theta_0} \left(\int_0^T \beta(X^i(t))^2 dt \right)^2 \right. \\
&\qquad \left. + \frac{8\overline{\mu}^4}{(\underline{\omega^2})^2} + 8(\overline{\omega^2})^2 \mathbb{E}_{\theta_0} \left(\int_0^T \beta(X^i(t)) dW^i(t) \right)^4 \right\} \\
&\leq C_\varepsilon + \frac{3}{4} \left\{ \frac{\overline{9\mu}^4}{(\underline{\omega^2})^2} + 8C^4 (\overline{\omega^2})^2 T \int_0^T \mathbb{E}_{\theta_0} \left(1 + \left| X^i(t) \right|^{8\gamma} \right) dt \right. \\
&\qquad \left. + 64 C_4 C^4 (\overline{\omega^2})^2 T \int_0^T \mathbb{E}_{\theta_0} \left(1 + \left| X^i(t) \right|^{8\gamma} \right) dt \right\} < \infty,
\end{aligned}$$

where C_4 is a nonnegative constant due to BDG inequality. In the previous inequalities, we systematically used Jensen inequality and the following facts:

$$\begin{aligned}
(a+b+c)^2 &\leq 3(a^2+b^2+c^2), \text{ for all } a,b,c \in \mathbb{R}; \\
|a+b|^r &\leq 2^{r-1}(|a|^r + |b|^r), \text{ for all } a,b \in \mathbb{R} \text{ and } r > 1; \\
\int_0^T \mathbb{E}_{\theta_0} \left| X^i(t) \right|^{2k} dt &< \infty, \text{ for all } \ k \geq 1 \text{ (for th proof see Appendix A. (p.104))}. \quad (4.9)
\end{aligned}$$

For nonnegative integers $\alpha_1, \cdots, \alpha_d$, we denote d-index $\alpha = (\alpha_1, \cdots, \alpha_d)$, $|\alpha| = \alpha_1 + \cdots + \alpha_d$ and

$$D^\alpha f(\theta) = \frac{\partial^{\alpha_1}}{\partial \theta_1^{\alpha_1}} \cdots \frac{\partial^{\alpha_d}}{\partial \theta_d^{\alpha_d}} f(\theta).$$

We introduce the following matrix as an alternative of Fisher information matrix:

$$\left(\mathcal{I}_X^{(\varepsilon)}(\theta_0) \right)_{k,j} = \mathbb{E}_{\theta_0} \left\{ \frac{\partial}{\partial \theta_k} \log \Lambda^{(\varepsilon)}(X^1, \theta) \frac{\partial}{\partial \theta_j} \log \Lambda^{(\varepsilon)}(X^1, \theta)|_{\theta=\theta_0} \right\}.$$

Proposition 4.1.5. *Let $\gamma_t(X^1, \theta)$ be three times continuously differentiable with respect to θ such that*

$$\mathbb{E}_{\theta_0} \left(\int_0^T (D^\alpha \gamma_t(X^1, \theta))^2 dt \right)^2 < \infty, \text{ for all } |\alpha| \leq 3.$$

The following statements hold true:

(i) *The matrix* $\mathcal{I}_X^{(\varepsilon)}(\theta_0)$ *is finite.*

(ii) *There is some* $r_0 > 0$ *and random function* $H(X^1, \theta_0)$ *depending only on* θ_0 *such that*

$$\sup_{\theta \in \overline{B_r(\theta_0)}} \left| D^2 \log \Lambda^{(\varepsilon)}(X^1, \theta_0) - D^2 \log \Lambda^{(\varepsilon)}(X^1, \theta) \right| \leq r H(X^1, \theta_0), \quad \textit{for } 0 < r < r_0$$

with $\mathbb{E}_{\theta_0} H(X^1, \theta_0) < \infty$.

Theorem 4.1.6. *Let the assumptions of Theorem 4.1.4 and Proposition 4.1.5 be satisfied. Assume further that the matrix* $\mathcal{I}_X^{(\varepsilon)}(\theta_0)$ *is nonsingular. Then,*

$$\sqrt{N}\left(\widehat{\theta}_N^{(\varepsilon)} - \theta_0\right) \overset{\mathcal{L}}{\Longrightarrow} \mathcal{N}\left(0, \mathcal{I}_X^{(\varepsilon)}(\theta_0)^{-1}\right).$$

Proof. To prove Theorem 4.1.6, we follow Schervish. Let $X = (X^1, \cdots, X^N)$ and set

$$l_\theta^{(\varepsilon)}(X) = \frac{1}{N} \sum_{i=1}^{N} \log \Lambda^{(\varepsilon)}(X^i, \theta).$$

The jth coordinate of the gradient of $l_\theta^{(\varepsilon)\prime}(X)$ is given by $(\sum_{i=1}^{N} \partial/\partial\theta_j \log \Lambda^{(\varepsilon)}(X^i, \theta))/N$. Since $\theta_0 \in \dot{\Theta}$, there is an open neighborhood $B(\theta_0, \delta) \subset \dot{\Theta}$. From the convergence of $\widehat{\theta}_N^{(\varepsilon)}$ to θ_0, it follows that $Z_N \chi_{\dot{\Theta}^c}(\widehat{\theta}_N^{(\varepsilon)}) = o_p(1/\sqrt{N})$ as $N \to \infty$ for every sequence $\{Z_N\}_{N=1}^{\infty}$ of random variables. In fact, for every $r_N > 0$, we have

$$\begin{aligned} \mathbb{P}_{\theta_0}\left(\left| Z_N \chi_{\dot{\Theta}^c}(\widehat{\theta}_N^{(\varepsilon)})/r_N \right| > \varepsilon'\right) &= \mathbb{P}_{\theta_0}\left(|Z_N/r_N| > \varepsilon' \,,\ \widehat{\theta}_N^{(\varepsilon)} \notin \dot{\Theta}\right) \\ &\leq \mathbb{P}_{\theta_0}\left(\widehat{\theta}_N^{(\varepsilon)} \notin B(\theta_0, \delta)\right) \longrightarrow 0, \text{ as } N \longrightarrow \infty, \text{ for all } \varepsilon' > 0. \end{aligned}$$

Note that $l_{\widehat{\theta}_N^{(\varepsilon)}}^{(\varepsilon)\,\prime}(X) = 0$, for $\widehat{\theta}_N^{(\varepsilon)} \in \dot{\Theta}$. Thus

$$l_{\widehat{\theta}_N^{(\varepsilon)}}^{(\varepsilon)\,\prime}(X) = l_{\widehat{\theta}_N^{(\varepsilon)}}^{(\varepsilon)\,\prime}(X) \chi_{\dot{\Theta}^c}(\widehat{\theta}_N^{(\varepsilon)}) = o_p(1/\sqrt{N})$$

By using one-term Taylor expansion of each coordinate $l_{\widehat{\theta}_N^{(\varepsilon)}}^{(\varepsilon)\,\prime}(X)$ around θ_0, we obtain

$$l_{\theta_0}^{(\varepsilon)\prime}(X) + \left(\left(\frac{\partial^2}{\partial\theta_k \partial\theta_j} l_\theta^{(\varepsilon)}(X)|_{\theta=\theta_\varepsilon^*}\right)\right)(\widehat{\theta}_N^{(\varepsilon)} - \theta_0) = o_p(1/\sqrt{N}), \tag{4.10}$$

where each $\theta_{\varepsilon,j}^*$ is a convex combination of $\widehat{\theta}_{N,j}^{(\varepsilon)}$ and $\theta_{0,j}$. Since $\widehat{\theta}_N^{(\varepsilon)} \overset{\mathbb{P}_{\theta_0}}{\Longrightarrow} \theta_0$, $\widehat{\theta}_{N,j}^{(\varepsilon)} \to \theta_{0,j}$ as well for each j under $\mathbb{P}_{\theta_0}$. Set $B_N^{(\varepsilon)}$ equal to the matrix in (4.10). Then

$$l_{\theta_0}^{(\varepsilon)\prime}(X) + B_N^{(\varepsilon)}(\widehat{\theta}_N^{(\varepsilon)} - \theta_0) = o_p(1/\sqrt{N}). \tag{4.11}$$

We have $\mathbb{E}_{\theta_0}\Lambda^{(\varepsilon)}(X^i,\theta_0) = 1$. By passing the derivatives under the integral sign in the previous equation, we get $0 = \mathbb{E}_{\theta_0}\frac{\partial}{\partial\theta_j}\Lambda^{(\varepsilon)}(X^i,\theta_0)$, hence $\mathbb{E}_{\theta_0}\left(l_{\theta_0}^{(\varepsilon)'}(X)\right) = 0$. Similarly, the conditional covariance matrix given $\theta = \theta_0$ of $l_{\theta_0}^{(\varepsilon)}(X)$ is $\mathcal{I}_X^{(\varepsilon)}(\theta_0)$. The multivariate central limit theorem (see, [101, Theorem B.99]) yields $\sqrt{N}l_{\theta_0}^{(\varepsilon)'}(X) \overset{\mathcal{L}}{\Longrightarrow} \mathcal{N}\left(0, \mathcal{I}_X^{(\varepsilon)}(\theta_0)\right)$ (since $\mathcal{I}_X^{(\varepsilon)}(\theta_0)$ is finite by **(i)** in Proposition 4.1.5. Hence, by Prohorov's theorem (see, e.g., [113, p.8]) and (4.11), we have

$$\sqrt{N}B_N^{(\varepsilon)}(\widehat{\theta}_N^{(\varepsilon)} - \theta_0) = O_p(1). \tag{4.12}$$

Note that $\left(B_N^{(\varepsilon)}\right)_{k,j} = \left(\sum_{i=1}^{N}\frac{\partial^2}{\partial\theta_k\partial\theta_j}\log\Lambda^{(\varepsilon)}(X^i,\theta)|_{\theta=\theta}\right)/N + \Delta_N$, with $|\Delta_N| \leq r\sum_{i=1}^{N}H(X^i,\theta_0)/N$ (this is justified by **(ii)** in Proposition 4.1.5, when $\|\theta_\varepsilon^* - \theta_0\| < r$). The weak law of large numbers yields $\sum_{i=1}^{N}H(X^i,\theta_0)/N \overset{\mathbb{P}_{\theta_0}}{\Longrightarrow} \mathbb{E}_{\theta_0}H(X^1,\theta_0)$. Let $\varepsilon' > 0$ and choose r to be small enough so that $r\mathbb{E}_{\theta_0}H(X^1,\theta_0) < \varepsilon'/2$. Then

$$\begin{aligned}
\mathbb{P}_{\theta_0}\left(|\Delta_N| > \varepsilon'\right) &\leq \mathbb{P}_{\theta_0}\left(\frac{r}{N}\sum_{i=1}^{N}H(X^i,\theta_0) > \varepsilon'\right) + \mathbb{P}_{\theta_0}\left(\|\theta_\varepsilon^* - \theta_0\| \geq r\right) \\
&\leq \mathbb{P}_{\theta_0}\left(\frac{1}{N}\sum_{i=1}^{N}H(X^i,\theta_0) - \mathbb{E}_{\theta_0}H(X^1,\theta_0) > \frac{\varepsilon'}{2r}\right) \\
&\qquad + \mathbb{P}_{\theta_0}\left(\left\|\widehat{\theta}_N^{(\varepsilon)} - \theta_0\right\| \geq r\right) \longrightarrow 0 \text{ as } N \longrightarrow \infty.
\end{aligned}$$

It follows that $\Delta_N = o_p(1)$, hence $B_N^{(\varepsilon)} \overset{\mathbb{P}_{\theta_0}}{\Longrightarrow} -\mathcal{I}_X^{(\varepsilon)}(\theta_0)$, and $B_N^{(\varepsilon)} = O_p(1)$ but $B_N^{(\varepsilon)} \neq o_p(1)$. As result, (4.12) gives $\sqrt{N}(\widehat{\theta}_N^{(\varepsilon)} - \theta_0) = O_p(1)$. Now, write$B_N^{(\varepsilon)} = -\mathcal{I}_X^{(\varepsilon)}(\theta_0) + C_N$, where $C_N = o_p(1)$. Then, by using the usual operations on "big-oh" and "small-oh", we have $C_N\left(\widehat{\theta}_N^{(\varepsilon)} - \theta_0\right) = o_p(1/\sqrt{N})$. Substituting $B_N^{(\varepsilon)}$ by its value in (4.11), we obtain

$$o_p(1/\sqrt{N}) = l_{\theta_0}^{(\varepsilon)'}(X) - \mathcal{I}_X^{(\varepsilon)}(\theta_0)(\widehat{\theta}_N^{(\varepsilon)} - \theta_0) + C_N(\widehat{\theta}_N^{(\varepsilon)} - \theta_0),$$

which can be rewritten as

$$\sqrt{N}l_{\theta_0}^{(\varepsilon)'}(X) - \mathcal{I}_X^{(\varepsilon)}(\theta_0)\sqrt{N}(\widehat{\theta}_N^{(\varepsilon)} - \theta_0) = o_p(1).$$

Finally, by continuous mapping theorem and Slutsky's lemma, we have

$$\begin{aligned}
\sqrt{N}(\widehat{\theta}_N^{(\varepsilon)} - \theta_0) &= \mathcal{I}_X^{(\varepsilon)}(\theta_0)^{-1}\left[\sqrt{N}l_{\theta_0}^{(\varepsilon)'}(X) - o_p(1)\right] \\
&\overset{\mathcal{L}}{\Longrightarrow} \mathcal{N}\left(0, \mathcal{I}_X^{(\varepsilon)}(\theta_0)^{-1}\right) \quad \text{under } \mathbb{P}_{\theta_0},
\end{aligned}$$

and the proof is complete. □

Remark. Note that the matrix $\mathcal{I}_X^{(\varepsilon)}(\theta_0)$ with small ε can be seen as an approximation of the Fisher information matrix $\mathcal{I}_X(\theta_0)$. As result, if $\mathcal{I}_X(\theta_0)$ is nonsingular, we may choose $\varepsilon > 0$ so that $\mathcal{I}_X^{(\varepsilon)}(\theta_0)$ is nonsingular as well.

4.1.4 Applications and Examples

This section is devoted to the study of a particular cases in which all basic assumptions are met. The results are illustrated by various examples, such as Wright-Fisher diffusion [61, p. 176] and Hyperbolic diffusion [13, p. 47].

Let $\beta_u(\varphi)$ be a nonnegative deterministic (continuous) function such that

$$\begin{aligned}\left|\beta(X^i(s),\varphi,\varphi_0)\right| &\le \beta_s(\varphi),\ \text{for all } s,\varphi\\ \text{and}\quad \int_0^T \beta_u(\varphi)^2 du &< \infty,\ \text{for all } \varphi\end{aligned}$$

Let us consider the following conditions:

$\mathbf{D}_1$: For any $\theta \in \Theta$, $\exists N_\theta$ an open set such that $\theta \in N_\theta$ and there exist $\delta > 0$, $C > 0$ and density function $h(\varphi)$ w.r.t ν so that

$$|g(\varphi,\theta_1) - g(\varphi,\theta_2)| \le Ch(\varphi)\left\|\theta_1 - \theta_2\right\|^\delta,\ \text{for all}\theta_1,\theta_2 \in N_\theta. \tag{4.13}$$

$\mathbf{D}_2$: For $f = h(\cdot)$ or $g(\cdot,\theta)$, we have

$$\int e^{\lambda\int_0^T \beta_u(\varphi)^2 du} f(\varphi)d\nu(\varphi) < \infty,\ \text{for all } \lambda > 0. \tag{4.14}$$

Proposition 4.1.7. *Let the conditions $\boldsymbol{D}_1$ and $\boldsymbol{D}_2$ be fulfilled. We have*

(i) $\mathbb{E}_{\theta_0}\int_0^{T_i} \gamma_s^2(X^i,\theta)ds < \infty,$ *for all* $i = 1,\cdots,N$.

(ii) *For each i, the function $\Lambda^{(\varepsilon)}(X^i,\theta)$ is continuous in θ $\mathbb{P}_{\theta_0}-a.s.$*

Example 4.1.2. Consider the hyporbolic diffusion $X = X^1$ with dynamics:

$$dX_t = \phi_1\frac{\phi_2 X_t}{\sqrt{1+\phi_2^2X_t^2}}dt + dW_t,\ X_0 = x \in \mathbb{R},$$

where ϕ_1 and ϕ_2 are two independent random variables with densities that satisfy $\mathbf{D}_1$ and ϕ_1 has compact support.

Example 4.1.3. Consider Wright-Fisher diffusion $X = X^1$ with dynamics:

$$dX_t = \phi X_t(1 - X_t)dt + \sqrt{X_t(1 - X_t)}dW_t,\ X_0 = x \in \mathbb{R},$$

where ϕ is a random variable with density having compact support and satisfies $\mathbf{D}_1$.

Example 4.1.4. Consider Double-Well potential diffusion $X = X^1$ with dynamics:

$$dX_t = \left(\phi_1 X_t - \phi_2 X_t^3\right) dt + dW_t,\ X_0 = x \in \mathbb{R},$$

where ϕ_1 and ϕ_2 are normally distributed.

Example 4.1.5. Consider the diffusion process $X = X^1$ with dynamics:

$$dX_t = \phi X_t dt + \sqrt{1 + X_t^2}dW_t,\ X_0 = x \in \mathbb{R},$$

where ϕ is normally distributed.

Remark. Many densities such as, Normal, Cauchy, Logistic and Gamma distributions satisfy $\mathbf{D}_1$. Hence, by truncation procedure, we can derive from them densities with compact support satisfying $\mathbf{D}_1$ & $\mathbf{D}_2$. It is worth to mention that cases where density $g(\varphi, \theta)$ has unbounded support are less important (since data can always be mapped monotonically to $[0, 1]$; and densities with unbounded support occur less often in practice). In contrast, the densities with compact support are easy to handle numerically.

4.1.5 Implementation Issues and Numerical Applications

For the implementation issue, the contrast functions are either explicitly computed as in Table 4.1 or approximated as in Table 4.2 and Table 4.3 by means of quadrature rules (e.g., [42]). More precisely, we use the following quadrature formula for a nonlinear cases. $\displaystyle\int_{-\infty}^{+\infty} h(u)e^{-u^2}du \simeq \sum_{r=1}^{R} h(z_r)w_r$, where $z_r = r-$ th zero of the the Hermite polynomial $H_R(u)$ with degree R. Three examples from previous section are numerically implemented later for different values of ε, N and R.

In practice, one rather disposes of discrete observations on the time interval $[0, T]$. Suppose that subjects X^i, $i = 1, 2, \cdots, N$ are discretly observed, only for simplicity, at equidistant points t_j, $j = 1, 2, \cdots, n$ of $[0, T]$. Then we discretize the integrals defining the conditional likelihood $L_T(X^i, \varphi, \varphi_0)$, that is,

$$\int_0^T \underbrace{H_2^+(X^i(s))^2 \left[H_1(X^i(s), \varphi) - H_1(X^i(s), \varphi_0)\right]}_{= A^i(s)} dX^i(s)$$

$$\simeq \sum_{j=1}^n A^i(t_{j-1})(X^i(t_j) - X^i(t_{j-1}))$$

$$\int_0^T \underbrace{H_2^+(X^i(s))^2 \left[H_1^2(X^i(s), \varphi) - H_1^2(X^i(s), \varphi_0)\right]}_{= B^i(s)} ds$$

$$\simeq \sum_{j=1}^n B^i(t_{j-1})(t_j - t_{j-1}).$$

Hence, we can get an approximate contrast function when an explicit one is not available. For the simulation studies, 100 dataest are generated for different numbers of subjects; $N = 50$,$N = 100$, $N = 1000$ with horizon time $T = 5$ and for different values of ε; $\varepsilon = 0.02$, $\varepsilon = 0.002$ and $\varepsilon = 0$ when the last value corresponds to the true maximum likelihood parameters estimation. The dataset are simulated as follows: we begin by drawing the random effect, then for each random effect drained value, the sample paths are simulated using Milstein scheme, at two different numbers ($n = 2^{10}$ or 2^8) of observations points very close in the fixed interval of time $[0, 5]$. The model parameters are estimated using gaussian density as random effects distribution.

Looking at Tables 4.1-4.3, they show that the estimations of the parameters are generally much closer to their true values; there is a considerable improvement as the value of ε decreases. The results computed from 100 dataset are excellent, even both N and R are not too large. For the Example 4.1.5, Table 4.2 shows that the estimation are much closer to their true values rather than those found in [25, Table 4]. For the nonlinear case (Example 4.1.2), Table 4.3 shows that the estimations are very satisfactory, and the accuracy can be improved by increasing only N. Compared to N (number of subjects), increasing R have no significant impact on the quality of the estimators and consume more time. Also the histograms in Figure 4.1 and Figure 4.2 reveal the asymptotic normality property of the estimators.

True parameter values	$\varepsilon = 0.02$	$\varepsilon = 0.002$	$\varepsilon = 0$
$N = 50$	**Mean (Std. dev.)**	**Mean (Std. dev.)**	**Mean (Std. dev.)**
$\mu_1 = -1$	-0.9394(0.4112)	-1.0127(0.3836)	-1.0048(0.2623)
$\sigma_1^2 = 0.81$	0.7143(0.4266)	0.8242(0.3839)	0.7928(0.3780)
$\mu_2 = 2$	1.9955(0.5589)	1.9040(0.4060)	2.0011(0.3966)
$\sigma_2^2 = 0.42$	0.3440(0.4585)	0.2610(0.2988)	0.3478(0.2786)
$N = 100$			
$\mu_1 = -1$	-0.9353(0.4421)	-0.9044(0.5223)	-0.9488(0.2232)
$\sigma_1^2 = 0.81$	0.8198(0.3397)	0.8431(0.2534)	0.7814(0.3039)
$\mu_2 = 2$	1.9263(0.3522)	1.8883(0.3751)	2.0569(0.2949)
$\sigma_2^2 = 0.42$	0.2718(0.2784)	0.3580(0.3059)	0.4066(0.2730)

Table 4.1. **Example 4.1.4:** $dX^i(t) = \left(\phi_{1,i}X^i(t) - \phi_{2,i}X^i(t)^3\right) dt + dW^i(t)$, $\phi_{s,i} \sim \mathcal{N}\left(\mu_s, \sigma_s^2\right)$, $s = 1, 2$. Empirical mean and Std. dev. (in brackets) of estimators $\widehat{\mu_1}$, $\widehat{\mu_2}$, $\widehat{\sigma_1^2}$, $\widehat{\sigma_2^2}$ are computed from 50 repeated simulated data sets for different values of (N, ε).

True parameter values	$\varepsilon = \mathbf{0.02}$	$\varepsilon = \mathbf{0.002}$	$\varepsilon = \mathbf{0}$
$\boldsymbol{N} = \mathbf{50}, \boldsymbol{R} = \mathbf{50}$	**Mean (Std. dev.)**	**Mean (Std. dev.)**	**Mean (Std. dev.)**
$\mu = -1$	-1.0251(0.2108)	-0.9703(0.2039)	-1.0145(0.2214)
$\sigma^2 = 1$	0.9659(0.3407)	0.9639(0.3819)	1.0007(0.3524)
$\mu = 5$	4.9955(0.1630)	4.9995(0.1761)	4.9879(0.1732)
$\sigma^2 = 1$	0.9690(0.2378)	0.9743(0.2443)	0.9784(0.2421)
$\boldsymbol{N} = \mathbf{100}, \boldsymbol{R} = \mathbf{100}$			
$\mu = -1$	-1.0309(0.1493)	-1.0234(0.1623)	-1.0036(0.1494)
$\sigma^2 = 1$	0.9574(0.2587)	1.0303(0.2308)	0.9692(0.2762)
$\mu = 5$	5.0081(0.1087)	5.0102(0.1093)	5.0153(0.1051)
$\sigma^2 = 1$	0.9684(0.1876)	0.9897(0.1807)	0.9665(0.1871)

Table 4.2. **Example 4.1.5:** $dX^i(t) = \phi_i X^i(t)dt + \sqrt{1 + X^i(t)^2}dW^i(t)$, $\phi_i \sim \mathcal{N}(\mu, \sigma^2)$. Empirical mean and Std. dev. (in brackets) of estimators $\widehat{\mu}$ and $\widehat{\sigma^2}$ are computed from 100 repeated simulated data sets for different values of (N, R, ε) and different parameter values.

True parameter values	$\varepsilon = 0.02$	$\varepsilon = 0.002$	$\varepsilon = 0$
$N = 10^3,\ R = 50$	**Mean (Std. dev.)**	**Mean (Std. dev.)**	**Mean (Std. dev.)**
$\mu = -1$	-1.0335 (0.0743)	-1.0055 (0.0706)	-1.0034 (0.0730)
$\sigma^2 = 1$	1.0353 (0.1481)	1.0040 (0.1367)	1.0063 (0.1444)
$N = 10^3,\ R = 100$			
$\mu = -1$	-1.0288 (0.0757)	-1.0062 (0.0715)	-0.9993 (0.0705)
$\sigma^2 = 1$	1.0324 (0.1643)	1.0131 (0.1418)	0.9967 (0.1448)

Table 4.3. **Example 4.1.2:** $dX^i(t) = \dfrac{\phi_i X^i(t)}{\sqrt{1+\phi_i^2 X^i(t)^2}} dt + dW^i(t)$, $\phi_i \sim \mathcal{N}\left(\mu, \sigma^2\right)$. Empirical mean and Std. dev. (in brackets) of estimators $\widehat{\mu}$ and $\widehat{\sigma^2}$are computed from hundreds of repeated simulated data sets for different values of (N, R, ε) and different parameter values.

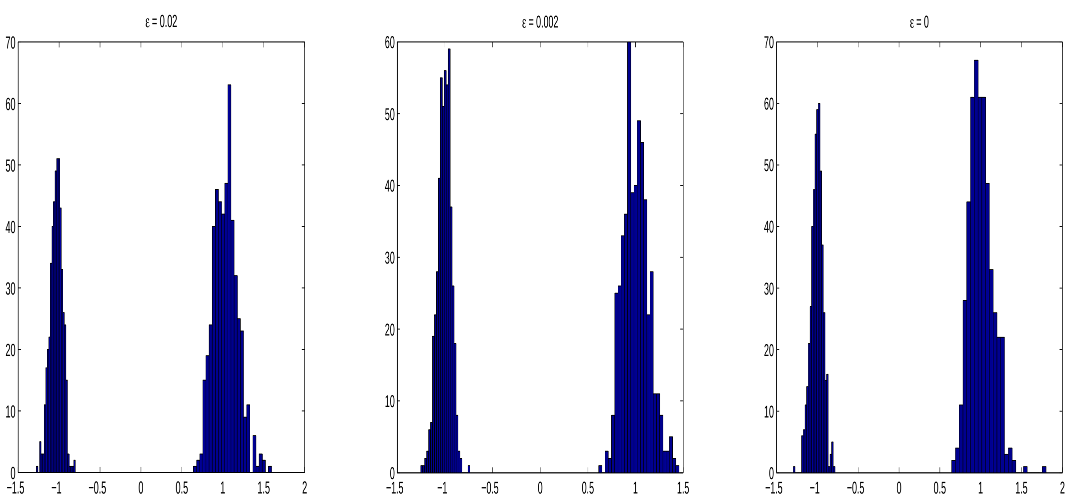

Figure 4.1. **Example 4.1.2:** Hyperbolic diffusion process, histograms of $\widehat{\mu}(\varepsilon)$ (on the left) and $\widehat{\sigma^2}(\varepsilon)$ (on the right), for different values of $\varepsilon = 0.02, 0.002, 0$ and fixed parameters: $(\mu, \sigma^2) = (-1, 1)$ and $(T, N, n, R) = (5, 1000, 2^8, 50)$.

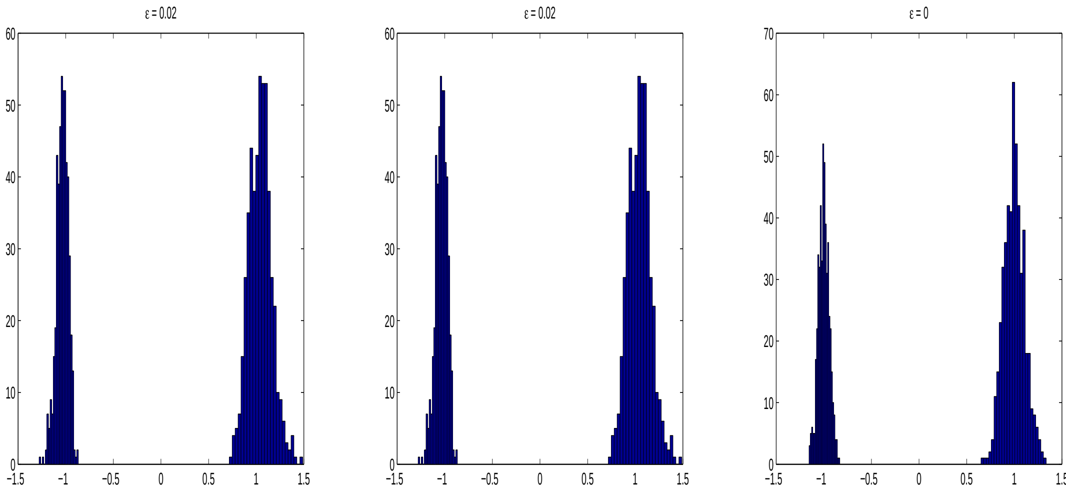

Figure 4.2. **Example 4.1.2:** Hyperbolic diffusion process, histograms of $\widehat{\mu}(\varepsilon)$ (on the left) and $\widehat{\sigma^2}(\varepsilon)$ (on the right), for different values of $\varepsilon = 0.02, 0.002, 0$ and fixed parameters: $(\mu, \sigma^2) = (-1, 1)$ and $(T, N, n, R) = (5, 1500, 2^8, 100)$.

4.2 NEW EXPANSION OF THE LIKELIHOOD FUNCTION

We adopt notations introduced in the previous section. Set

$$\alpha_n^i(t_1, t_2, \ldots, t_n, \theta) = \int \beta(X^i(t_1), \varphi, \varphi_0) \cdots \beta(X^i(t_n), \varphi, \varphi_0) g(\varphi, \theta) d\nu(\varphi),$$

where $\beta(x, \varphi, \varphi_0) = H_2^+(x)\left(H_1(x, \varphi) - H_1(x, \varphi_0)\right)$, for all $x \in \mathbb{R}$ and $\varphi, \varphi_0 \in \mathbb{R}^p$. For each subject X^i, we denote by $J_n\left(\beta_i^{\otimes n}(\theta)\right)$, $n \geq 0$ the multiple integral defined as

$$\begin{aligned}
J_0\left(\beta_i^{\otimes 0}(\theta)\right) &= 1 \\
J_1\left(\beta_i^{\otimes 1}(\theta)\right) &= \int_0^T \alpha_1^i(t_1, \theta) dW^i(t_1) \\
J_2\left(\beta_i^{\otimes 2}(\theta)\right) &= \int_0^T \int_0^{t_1} \alpha_2^i(t_1, t_2, \theta) dW^i(t_2) dW^i(t_1) \\
\vdots \qquad & \vdots \qquad \vdots \qquad \qquad \vdots \\
J_n\left(\beta_i^{\otimes n}(\theta)\right) &= \int_0^T \int_0^{t_1} \cdots \int_0^{t_{n-1}} \alpha_n^i(t_1, t_2, \ldots, t_n, \theta) dW^i(t_n) \cdots dW^i(t_2) dW^i(t_1).
\end{aligned}$$

With these notations and argumenets, we are ready to expand the likelihood functions $\Lambda_i(X^i, \theta)$ as a series of $J_n\left(\beta_i^{\otimes n}(\theta)\right)$.

Proposition 4.2.1. *Let the conditions* $\boldsymbol{A}_1$- $\boldsymbol{A}_3$ *be satisfied and let* $\varphi, \varphi_0 \in \mathbb{R}^p$. *Then the individual density functions* $\Lambda_i(X^i, \theta)$ *defined in Proposition 4.1.1 can be rewritten as*

$$\Lambda_i(X^i, \theta) = \sum_{n=0}^{\infty} J_n\left(\beta_i^{\otimes n}(\theta)\right), \tag{4.15}$$

provided that $\int_{[0,T]^n} \mathbb{E}\alpha_n^i(t_1, t_2, \ldots, t_n, \theta)^2 dt_1 \cdots dt_n < \infty$.

Proof. For a fixed subject X^i, $i = 1, \cdots, N$, we simplify notations by setting $X^i = X$, $W^i = W$, $\beta_t = \beta(X^i(t), \varphi, \varphi_0)$ and $L = \left(L_t = L_T(X^i, \varphi, \varphi_0)|_{T=t}\ ,\ t \leq T\right)$. Clearly, $L_t = \exp\left(\int_0^t \beta_s dW_s - \frac{1}{2}\int_0^t \beta_s^2 ds\right)$ for $t \geq 0$. Thus L is a nonnegative supermartingale permitting the following representation

$$L_t = 1 + \int_0^t \beta_s L_s dW_s, \quad \text{for all } t \leq T. \tag{4.16}$$

Note that (4.16) follows immediately from Itô formula. Applying (4.16) recursively, we have

$$\begin{aligned}
L_T &= 1 + \int_0^T \beta_{t_1} L_{t_1} dW_{t_1} \\
&= 1 + \int_0^T \beta_{t_1}\left[1 + \int_0^{t_1} \beta_{t_2} L_{t_2} dW_{t_2}\right] dW_{t_1} \\
&= 1 + \int_0^T \beta_{t_1} dW_{t_1} + \int_0^T \int_0^{t_1} \beta_{t_1}\beta_{t_2}\left[1 + \int_0^{t_2} \beta_{t_3} L_{t_3} dW_{t_3}\right] dW_{t_2} dW_{t_1} \\
&\vdots \qquad \vdots \qquad \vdots \qquad \vdots \qquad \vdots \\
&= 1 + \int_0^T \beta_{t_1} dW_{t_1} + \cdots + \int_0^T \int_0^{t_1} \cdots \int_0^{t_{n-1}} \beta_{t_1}\beta_{t_2}\cdots\beta_{t_n} dW_{t_n} \cdots dW_{t_2} dW_{t_1} + \cdots \\
&= \sum_{n=0}^{\infty} J_n(\beta^{\otimes n}),
\end{aligned} \tag{4.17}$$

where $\beta^{\otimes n}$ is the tensor product defined by $\beta^{\otimes n}(t_1, \cdots, t_n) = \prod_{i=1}^{n} \beta_{t_i}$ and $J_n(f(\cdot))$ is the multiple stochastic integral defined by

$$J_n(f(\cdot)) = \int_0^T \int_0^{t_1} \cdots \int_0^{t_{n-1}} f(t_1, t_2, \cdots, t_n) dW_{t_n} \cdots dW_{t_2} dW_{t_1}$$

for $n \geq 1$ and $J_0(f(\cdot)) = 1$. Substituting (4.17) into the expression of the individual densities, then using the condition presented in **(ii)** and Fubini theorem for stochastic integrals (see, [73, Theorem 5.15]) to get the desired result. The condition presented in Proposition 4.2.1 holds true under the assumptions $\mathbf{A}_4$-$\mathbf{A}_5$ below. $\square$

4.2.1 Strong Consistency of the MLE

The following assumptions are needed to prove the consistency of the MLE.

A$_4$: There exist a nonnegative constants $M > 0$ and $\gamma > 0$ such that $|\beta(x, \varphi, \varphi_0)| \leq M(1 + |x|^\gamma)$, for all $x \in \mathbb{R}$ and $\varphi \in \mathbb{R}^p$.

A$_5$: There exists $M_2 > 0$ such that $\mathbb{E}_\theta \sup_{t \leq T} \left|X^i(t)\right|^{2k} \leq M_2^k$, for all k, $i = 1, \cdots, N$ and $\theta \in \Theta$.

A$_6$: For any $\theta \in \Theta$, $\exists N_\theta$ an open set such that $\theta \in N_\theta$ and there exist $\delta > 0$, $C > 0$ and density function $h(\varphi)$ w.r.t ν so that

$$|g(\varphi, \theta_1) - g(\varphi, \theta_2)| \leq Ch(\varphi) \left\|\theta_1 - \theta_2\right\|^\delta, \quad \forall \theta_1, \theta_2 \in N_\theta. \tag{4.18}$$

Instead of **A**$_6$ we may assume that $\mathbb{E}_\theta \sup_{t \leq T} \left|X^i(t)\right|^{2k} \leq M_1$, for all k, $i = 1, \cdots, N$ and $\theta \in \Theta$, for some fixed $M_1 > 0$. For $\theta \in \Theta$ set $\xi_\theta(x) = \displaystyle\int \beta(x, \varphi, \varphi_0) d\mu^\theta(\varphi)$ and $[\xi_\theta(x)] = \displaystyle\int \beta(x, \varphi, \varphi_0)^2 d\mu^\theta(\varphi)$ where $d\mu^\theta(\varphi) = g(\varphi, \theta) d\nu(\varphi)$. Under the previous assumptions and notations, we state the following results.

Proposition 4.2.2.

(i) *The random functions $J_n(\beta_i^{\otimes n}(\theta))$, $n \geq 1$ are continuous in θ, $\mathbb{P}_{\theta_0}$-a.s.*

(ii) *The individual likelihood function $\Lambda_i(X^i, \theta)$ is continuous in θ, $\mathbb{P}_{\theta_0}$-a.s.*

(iii) *For $\theta \in \Theta$ and $i = 1, \cdots, N$ $\mathbb{E}_{\theta_0} \displaystyle\int_0^T \left[\xi_\theta(X^i(s))\right] ds < \infty$.*

(iv) *For each θ and $i = 1, \cdots, N$, the process $\left(\displaystyle\int_0^t \xi_\theta(X^i(s)) dW^i(s); \quad t \geq 0\right)$ is a true martingale.*

(v) $\mathbb{E}_{\theta_0} \log \Lambda_i(X^i, \theta) > -\infty$, *for all $\theta \in \Theta$.*

(vi) $\mathbb{E}_{\theta_0} \log \Lambda_i(X^i, \theta) \leq 1$, *for all $\theta \in \Theta$.*

Proof. **(i)** We shall only focus on the properties of $\Lambda(X, \theta) = \Lambda_1(X^1, \theta)$ since the same procedures can be applied to other densities. We simplify notations by setting $X^i = X$, $W^i = W$, $\beta = \beta_1$, $dt^{\otimes n} = dt_n \cdots dt_1$, $dW^{\otimes n}(t) = dW(t_n) \cdots dW(t_1)$ and

$S_n = \{(t_1, \cdots, t_n) \in [0,T]^n \; ; \; 0 \le t_n \le t_{n-1} \le \cdots \le t_1 \le T\}$. Let N_θ be the open set such that $\theta \in N_\theta$ and (4.18) holds. Let $\theta_1, \theta_2 \in N_\theta$, $r > 0$ and set

$$\zeta_{\theta_1,\theta_2} = \int_{S_n} \left[\int \prod_{l=1}^{n} \beta(X(t_l), \varphi, \varphi_0) \left(g(\varphi, \theta_1) - g(\varphi, \theta_2) \right) d\nu(\varphi) \right] dW^{\otimes n}(t).$$

We shall verify the Kolmogorov-Chentsov criterion (2.3) with constants to be specified latter. Applying BDG and Jensen inequalities and Fubini theorem respectively, we claim that there exists $C_r > 0$ such that

$$\begin{aligned}
&\mathbb{E}_{\theta_0} \left| J_n\left(\beta^{\otimes n}(\theta_1)\right) - J_n\left(\beta^{\otimes n}(\theta_2)\right) \right|^{2r} = \mathbb{E}_{\theta_0} \zeta^{2r}_{\theta_1,\theta_2} \le C_r \mathbb{E}_{\theta_0} \langle \zeta_{\theta_1,\theta_2} \rangle^r \\
&\le T^{r-1} C_r \int_0^T \mathbb{E}_{\theta_0} \left\{ \int_{S_{n-1}} \left(\int \prod_{l=1}^{n} \beta(X(t_l), \varphi, \varphi_0) \left(g(\varphi, \theta_1) - g(\varphi, \theta_2) \right) d\nu(\varphi) \right) dW^{\otimes n-1}(t) \right\}^{2r} dt_1 \\
&\le (T^{r-1} C_r)^2 \int_0^T \int_0^{t_1} \mathbb{E}_{\theta_0} \left\{ \int_{S_{n-2}} \left(\int \prod_{l=1}^{n} \beta(X(t_l), \varphi, \varphi_0) \left(g(\varphi, \theta_1) - g(\varphi, \theta_2) \right) d\nu(\varphi) \right) dW^{\otimes n-2}(t) \right\}^{2r} dt_2 dt_1 \\
&\qquad \vdots \\
&\le (T^{r-1} C_r)^n \int_{S_n} \mathbb{E}_{\theta_0} \left\{ \int \prod_{l=1}^{n} \beta(X(t_l), \varphi, \varphi_0) \left(g(\varphi, \theta) - g(\varphi, \theta') \right) d\nu(\varphi) \right\}^{2r} dt^{\otimes n} \\
&\le (T^{r-1} C_r)^n \int_{S_n} \mathbb{E}_{\theta_0} \left\{ \int \prod_{l=1}^{n} \beta(X(t_l), \varphi, \varphi_0) C \left\| \theta_1 - \theta_2 \right\|^\delta h(\varphi) d\nu(\varphi) \right\}^{2r} dt^{\otimes n} \\
&\le (T^{r-1} C_r)^n C^{2r} \left\| \theta_1 - \theta_2 \right\|^{2r\delta} \int_{S_n} \left(\int \mathbb{E}_{\theta_0} \left\{ \prod_{l=1}^{n} \beta(X(t_l), \varphi, \varphi_0) \right\}^{2r} h(\varphi) d\nu(\varphi) \right) dt^{\otimes n} \\
&\le \frac{(T^{r-1} C_r)^n}{n!} C^{2r} \left\| \theta_1 - \theta_2 \right\|^{2r\delta} \int_{[0,T]^n} \left(\int \mathbb{E}_{\theta_0} \prod_{l=1}^{n} \beta(X(t_l), \varphi, \varphi_0)^{2r} h(\varphi) d\nu(\varphi) \right) dt^{\otimes n}
\end{aligned}$$

We can choose $r \ge \lceil \frac{d}{2\delta} \rceil$ and $\varepsilon = 2r\delta - d$, so that the Kolmogorov-Chentsov criterion holds. Hence, the continuity of $J_n(\beta_1^{\otimes n}(\theta))$ in θ, $\mathbb{P}_{\theta_0}$-a.s holds true.

(ii) Let $\theta \in \Theta$ and $r \ge 1$, we have $\Lambda(X, \theta) = \sum_{n \ge 0} J_n\left(\beta^{\otimes n}(\theta)\right) = \lim_{m \longrightarrow \infty} \Lambda^{(m)}(X, \theta)$ where $\Lambda^{(m)}(X, \theta) = \sum_{n=0}^{m} J_n\left(\beta^{\otimes n}(\theta)\right)$. By using Minkowski inequality, we obtain

$$\left\{\mathbb{E}_{\theta_0}\left|\Lambda^m(X,\theta_1)-\Lambda^m(X,\theta_2)\right|^{2r}\right\}^{1/2r} = \left\{\mathbb{E}_{\theta_0}\left|\sum_{n=1}^{m} J_n\left(\beta^{\otimes n}(\theta_1)-\beta^{\otimes n}(\theta_2)\right)\right|^{2r}\right\}^{1/2r}$$

$$\leq \left\{\mathbb{E}_{\theta_0}\left(\sum_{n=1}^{m}\left|J_n\left(\beta^{\otimes n}(\theta_1)-\beta^{\otimes n}(\theta_2)\right)\right|\right)^{2r}\right\}^{1/2r}$$

$$\leq \sum_{n=1}^{m}\left\{\mathbb{E}_{\theta_0}\left(\left|J_n\left(\beta^{\otimes n}(\theta_1)-\beta^{\otimes n}(\theta_2)\right)\right|\right)^{2r}\right\}^{1/2r} \leq \sum_{n=1}^{m}\left\{\mathbb{E}_{\theta_0}\zeta_{\theta_1,\theta_2}^{2r}\right\}^{1/2r}$$

$$\leq C\left\|\theta_1-\theta_2\right\|^{\delta}\sum_{n=1}^{m}\frac{(T^{r-1}C_r)}{\sqrt[2r]{n!}}\left\{\int_{[0,T]^n}\left(\int\mathbb{E}_{\theta_0}\prod_{l=1}^{n}\beta(X(t_l),\varphi,\varphi_0)^{2r}h(\varphi)d\nu(\varphi)\right)dt^{\otimes n}\right\}^{1/2r} \quad (4.19)$$

$$\leq C\left\|\theta_1-\theta_2\right\|^{\delta}\sum_{n=1}^{m}\frac{2^{n-1/2r}M^nT^{n/2r}(T^{r-1}C_r)^{n/2r}}{\sqrt[2r]{n!}}\left(1+\mathbb{E}_{\theta_0}\sup_{t\leq T}|X_t|^{2\gamma nr}\right)^{1/2r}$$

$$\leq C\left\|\theta_1-\theta_2\right\|^{\delta}\sum_{n=1}^{\infty}\frac{2^{n-1/2r}M^nT^{n/2}C_r^{n/2r}}{\sqrt[2r]{n!}}\left(1+\mathbb{E}_{\theta_0}\sup_{t\leq T}|X_t|^{2\gamma nr}\right)$$

$$\leq C\times C(r,T,M,M_2)\left\|\theta_1-\theta_2\right\|^{\delta},$$

where $C(r,T,M,M_2)=\sum_{n=1}^{\infty}\frac{\left[2MT^{1/2}C_r^{1/2r}\right]^n}{\sqrt[2r]{n!}}(1+M_2^{\gamma nr})<\infty$. Now, we are ready to apply Fatou lemma to the sequence $\left\{\left|\Lambda^m(X,\theta_1)-\Lambda^m(X,\theta_2)\right|^{2r}\right\}_{m=1}^{\infty}$ to get

$$\begin{aligned}\left\{\mathbb{E}_{\theta_0}\left|\Lambda(X,\theta_1)-\Lambda(X,\theta_2)\right|^{2r}\right\}^{1/2r} &\leq \liminf_{m\longrightarrow\infty}\left\{\mathbb{E}_{\theta_0}\left|\Lambda^m(X,\theta_1)-\Lambda^m(X,\theta_2)\right|^{2r}\right\}^{1/2r}\\ &\leq C\times C(r,T,M,M_2)\left\|\theta_1-\theta_2\right\|^{\delta}.\end{aligned}$$

Hence, $\mathbb{E}_{\theta_0}\left|\Lambda(X,\theta_1)-\Lambda(X,\theta_2)\right|^{2r}\leq C^{2r}\times C(r,T,M,M_2)^{2r}\left\|\theta_1-\theta_2\right\|^{2r\delta}$ for all $\theta_1,\theta_2\in N_\theta$. Choosing again $r\geq\lceil\frac{d}{2\delta}\rceil\vee 1$ and $\varepsilon=2r\delta-d$, so that the Kolmogorov-Chentsov criterion holds. Thus, $\Lambda(X,\theta)$ is continuous in θ, $\mathbb{P}_{\theta_0}$-a.s.

For the proof of the statements **(iii)-(iv)**, we use Fubini theorem

$$\mathbb{E}_{\theta_0}\int_0^T\left[\xi_\theta(X_s)\right]ds=\mathbb{E}_{\theta_0}\int_0^T\int\beta(X_s,\varphi,\varphi_0)^2d\mu^\theta(\varphi)ds$$

$$\leq\int_0^T\int\mathbb{E}_{\theta_0}\beta(X_s,\varphi,\varphi_0)^2d\mu^\theta(\varphi)ds\leq M^2\int_0^T\int\mathbb{E}_{\theta_0}\left(1+|X_s|^\gamma\right)^2d\mu^\theta(\varphi)ds$$

$$\leq 2M^2\int_0^T\mathbb{E}_{\theta_0}\left(1+\mathbb{E}_{\theta_0}|X_s|^{2\gamma}\right)ds\leq 2M^2T\left(1+\mathbb{E}_{\theta_0}\sup_{t\leq T}|X_t|^{2\gamma}\right)<\infty.$$

From **(iii)** it follows that $\mathbb{E}_{\theta_0} \int_0^T (\xi_\theta(X_s))^2 \, ds < \infty$. Thus the process
$\eta = \left(\int_0^t \xi_\theta(X_s) dW_s; \quad t \geq 0 \right)$ is a true martingale. **(v)** By virtue of Jensen inequality and the martingale property of η, we have

$$\begin{aligned} \mathbb{E}_{\theta_0} \log \Lambda(X, \theta) &\geq \mathbb{E}_{\theta_0} \log \int L_T(X, \varphi, \varphi_0) d\mu^\theta(\varphi) \\ &\geq \mathbb{E}_{\theta_0} \int \log L_T(X, \varphi, \varphi_0) d\mu^\theta(\varphi) \\ &\geq \mathbb{E}_{\theta_0} \eta - \frac{1}{2} \mathbb{E}_{\theta_0} \int_0^T [\xi_\theta(X_s)] \, ds > -\infty. \end{aligned}$$

(vi) follows immediately from the supermartingale property of
$L = (L_T(X, \varphi, \varphi_0) \ : \ T \geq 0)$ and the proof of Proposition 4.2.2 is complete. □

The most delicate task of this work is to show that $\Lambda_i(X^i, \theta)$ is continuous in θ, $\mathbb{P}_{\theta_0}$-a.s. The proof of **(iii)**-**(vi)** requires weaker assumptions than those already considered in $\mathbf{A}_4$-$\mathbf{A}_5$. In fact, we need only that

- There exists a nonnegative polynomial $\mathcal{P}(x, \varphi)$ of order m (which may depend on φ_0) such that $|\beta(x, \varphi, \varphi_0)| \leq |\mathcal{P}(x, \varphi)|$, for all x and φ.
- The random effects have all moments of oder k, $k \leq 2m$.
- For all $k > 0$, $\sup_{t \leq T} \mathbb{E}_{\theta_0} \left| X^i(t) \right|^k < \infty$.

Theorem 4.2.3. *Under the assumptions* $\boldsymbol{A}_1$-$\boldsymbol{A}_6$*, the MLE* $\widehat{\theta}_N$ *is strongly consistent, i.e.,*

$$\widehat{\theta}_N \overset{\mathbb{P}_{\theta_0}-a.s}{\Longrightarrow} \theta_0, \ \ as \ N \longrightarrow \infty$$

Proof. The previous results (Proposition 4.2.2) combined with the compactness of the parameter space Θ yield the strong consistency of the MLE (see, Theorems 3.0.4-3.0.5). □

The previous results can be extended to non homogeneous diffusions. In fact the individual densities will be given by

$$\Lambda_i(X^i, \theta) = \int L_T(X^i, \varphi, \varphi_0) d\mu^\theta(\varphi),$$

where

$$L_T(X^i, \varphi, \varphi_0) = e^{\int_0^T \beta(s, X^i(s), \varphi, \varphi_0) dW^i(s) - \frac{1}{2} \int_0^T \beta(s, X^i(s), \varphi, \varphi_0)^2 ds}$$

If $\beta(s, X^i(s), \varphi, \varphi_0)$ is deterministic, that is $\beta(s, X^i(s), \varphi, \varphi_0) = \mathcal{A}(s, \varphi, \varphi_0)$, then the expansion of $\Lambda_i(X^i, \theta)$ given in (4.15) is exactly its Wiener-Itô chaos expansion. Black-Derman-Toy model is an example of this type (see, e.g., [13]). In this case, we have also the following Proposition.

Proposition 4.2.4.

(i) *For $i = 1, \cdots, N$ and $n \in \mathbb{N}$, $J_n(\beta_i^{\otimes n}(\theta)) \in L^2(\mathbb{P}_{\theta_0})$.*

(ii) *If $\Lambda_i(X, \theta) \in L^2(\mathbb{P}_{\theta_0})$ and $\beta(s, X^i(s), \varphi, \varphi_0)$ is deterministic, then the series $\sum_{n\geq 0} \left\| J_n(\beta_i^{\otimes n}(\theta)) \right\|^2_{L^2(\mathbb{P}_{\theta_0})}$ converges and*

$$\|\Lambda_i(X, \theta)\|^2_{L^2(\mathbb{P}_{\theta_0})} = \sum_{n\geq 0} \left\| J_n(\beta_i^{\otimes n}(\theta)) \right\|^2_{L^2(\mathbb{P}_{\theta_0})}.$$

Proof. **(i)** Let $\theta, \theta' \in \Theta$. Making use of the notations introduced in the proof of Proposition 4.2.2, we have

$$\begin{aligned}
\left\| J_n\left(\beta^{\otimes n}(\theta)\right) \right\|^2_{L^2(\mathbb{P}_{\theta'})} &\leq C_r^n \int_{S_n} \mathbb{E}_{\theta'} \left\{ \int \prod_{l=1}^{n} \beta(X(t_l), \varphi, \varphi_0) d\mu^\theta(\varphi) \right\}^2 dt^{\otimes n} \\
&\leq 2^{n-1} \frac{(TM^2 C_r)^n}{n!} \left(1 + \mathbb{E}_{\theta'} \sup_{t \leq T} |X_t|^{2\gamma n} \right) < \infty.
\end{aligned}$$

Therefore $J_n\left(\beta^{\otimes n}(\theta)\right) \in L^2(\mathbb{P}_{\theta'})$, for all (fixed) $\theta' \in \Theta$.
By induction (see the proof of Proposition 4.2.1), we can show that

$$\begin{aligned}
L_T(X, \varphi, \varphi_0) &= \sum_{n=0}^{m} J_n(\mathcal{A}(\cdot, \varphi, \varphi_0)^{\otimes n}) + \psi_{m+1}(\varphi, \varphi_0), \\
\text{where } \psi_{m+1}(\varphi, \varphi_0) &= \int_{S_{m+1}} \prod_{l=1}^{m+1} \mathcal{A}(t_l, \varphi, \varphi_0) L_{t_{m+1}}(X, \varphi, \varphi_0) dW^{\otimes m+1}(t).
\end{aligned}$$

By virtue of Itô isometry, simple computations lead to the following results

$$\begin{aligned}
\mathbb{E}_{\theta_0}\left(J_n\left(\mathcal{A}(\cdot, \varphi, \varphi_0)^{\otimes n} \psi_{m+1}(\varphi', \varphi_0)\right)\right) &= 0, \quad \forall n \leq m, \ \forall \varphi, \varphi' \\
\mathbb{E}_{\theta_0}\left(J_n\left(\mathcal{A}(\cdot, \varphi, \varphi_0)^{\otimes n}\right) J_k\left(\mathcal{A}(\cdot, \varphi', \varphi_0)^{\otimes k}\right)\right) &= 0, \quad \forall n \neq k, \ \forall \varphi, \varphi'.
\end{aligned}$$

Hence, for each φ, φ' we have

$$\begin{aligned}
&\mathbb{E}_{\theta_0}\left(L_T(X,\varphi,\varphi_0)L_T(X,\varphi',\varphi_0)\right)\\
&\qquad= \mathbb{E}_{\theta_0}\left\{\left(\sum_{n=0}^{m} J_n(\mathcal{A}(\cdot,\varphi,\varphi_0)^{\otimes n}) + \psi_{m+1}(\varphi,\varphi_0)\right)\right.\\
&\qquad\qquad \left.\times\left(\sum_{n=0}^{m} J_n(\mathcal{A}(\cdot,\varphi',\varphi_0)^{\otimes n}) + \psi_{m+1}(\varphi',\varphi_0)\right)\right\}\\
&= \sum_{n=0}^{m} \mathbb{E}_{\theta_0}\left\{J_n(\mathcal{A}(\cdot,\varphi,\varphi_0)^{\otimes n})J_n(\mathcal{A}(\cdot,\varphi',\varphi_0)^{\otimes n})\right\} + \mathbb{E}_{\theta_0}\left(\psi_{m+1}(\varphi,\varphi_0)\psi_{m+1}(\varphi',\varphi_0)\right).
\end{aligned}$$

Which implies that

$$\begin{aligned}
\mathbb{E}_{\theta_0}\Lambda(X,\theta)^2 &= \mathbb{E}_{\theta_0}\left(\int L_T(X,\varphi,\varphi_0)d\mu^\theta(\varphi)\right)^2\\
&= \mathbb{E}_{\theta_0}\int\int L_T(X,\varphi,\varphi_0)L_T(X,\varphi',\varphi_0)d\mu^\theta(\varphi)d\mu^\theta(\varphi')\\
&= \int\int \mathbb{E}_{\theta_0}\left(L_T(X,\varphi,\varphi_0)L_T(X,\varphi',\varphi_0)\right)d\mu^\theta(\varphi)d\mu^\theta(\varphi')\\
&= \sum_{n=0}^{m}\int\int \mathbb{E}_{\theta_0}\left\{J_n(\mathcal{A}(\cdot,\varphi,\varphi_0)^{\otimes n})J_n(\mathcal{A}(\cdot,\varphi',\varphi_0)^{\otimes n})\right\}d\mu^\theta(\varphi)d\mu^\theta(\varphi')\\
&\qquad + \int\int \mathbb{E}_{\theta_0}\left\{\psi_{m+1}(\varphi,\varphi_0)\psi_{m+1}(\varphi',\varphi_0)\right\}d\mu^\theta(\varphi)d\mu^\theta(\varphi')\\
&= \sum_{n=0}^{m}\mathbb{E}_{\theta_0}\left\{\int\int J_n(\mathcal{A}(\cdot,\varphi,\varphi_0)^{\otimes n})J_n(\mathcal{A}(\cdot,\varphi',\varphi_0)^{\otimes n})d\mu^\theta(\varphi)d\mu^\theta(\varphi')\right\}\\
&\qquad + \mathbb{E}_{\theta_0}\left\{\int\int \psi_{m+1}(\varphi,\varphi_0)\psi_{m+1}(\varphi',\varphi_0)d\mu^\theta(\varphi)d\mu^\theta(\varphi')\right\}\\
&= \sum_{n=0}^{m}\left\|J_n(\beta^{\otimes n}(\theta))\right\|^2_{L^2(\mathbb{P}_{\theta_0})} + \mathbb{E}_{\theta_0}\psi_{m+1}(\theta)^2,
\end{aligned}$$

where $\psi_{m+1}(\theta) := \int \psi_{m+1}(\varphi,\varphi_0)d\mu^\theta(\varphi)$. Since $\Lambda(X,\theta) \in L^2(\mathbb{P}_{\theta_0})$, then $\sum_{n\geq 0}\left\|J_n(\beta^{\otimes n}(\theta))\right\|^2_{L^2(\mathbb{P}_{\theta_0})} < \infty$ and $\{\psi_{m+1}(\theta)\}_{m=1}^{\infty}$ converges in $L^2(\mathbb{P}_{\theta_0})$.
Set $\psi(\theta) := \lim_{m\longrightarrow\infty}\psi_{m+1}(\theta)$. It remains to show that $\psi(\theta) = 0$. For that purpose, we use the same procedures as in the proof of Theorem 1.10 in [28]. For all $n \leq m$ and $f_n \in L^2([0,T]^n)$,

we have

$$\begin{aligned}
\mathbb{E}_{\theta_0}\left(J_n(f_n)\psi_{m+1}(\theta)\right) &= \mathbb{E}_{\theta_0}\left\{J_n(f_n)\int \psi_{m+1}(\varphi,\varphi_0)d\mu^\theta(\varphi)\right\} \\
&= \mathbb{E}_{\theta_0}\left\{\int J_n(f_n)\psi_{m+1}(\varphi,\varphi_0)d\mu^\theta(\varphi)\right\} \\
&= \int \mathbb{E}_{\theta_0}\left\{J_n(f_n)\psi_{m+1}(\varphi,\varphi_0)\right\}d\mu^\theta(\varphi) = 0.
\end{aligned}$$

The last equality can be ridely justified by Itô isometry.
Therefore $\mathbb{E}_{\theta_0}\left(J_n(f_n)\psi(\theta)\right) = 0$, for all $n \geq 0$ and $f_n \in L^2([0,T]^n)$. in particular, by Proposition 1.18 in [28], we have

$$\mathbb{E}_{\theta_0}\left(h_n\left(\frac{\int_0^T f(t)dW_t}{\|f\|}\right)\cdot\psi(\theta)\right) = 0.$$

Using the fact that x^n can be expressed as a linear combination of the Hermite polynomials $h_r(x)$, $0 \leq r \leq n$, we get $\mathbb{E}_{\theta_0}\left(\left(\int_0^T f(t)dW_t\right)^n\cdot\psi(\theta)\right) = 0$, for all $n \geq 0$, which implies again that

$$\mathbb{E}_{\theta_0}\left(e^{\int_0^T f(t)dW_t}\psi(\theta)\right) = \sum_{n=0}^{\infty}\frac{1}{n!}\mathbb{E}_{\theta_0}\left(\left(\int_0^T f(t)dW_t\right)^n\cdot\psi(\theta)\right) = 0.$$

Since the family $\left\{e^{\int_0^T f(t)dW_t}\ ;\ f \in L^2([0,T])\right\}$ is dense in $L^2(\mathbb{P}_{\theta_0})$ (see, [87, Lemma 4.3.2]), we conclude that $\psi(\theta) = 0$ and the proof is complete. □

Under the assumptions $\mathbf{A}_1$- $\mathbf{A}_6$, one can prove that the series given in (4.15) converges in $L^{2r}(\mathbb{P}_{\theta_0})$, but it may not be unique as expansion of the individual likelihood.

4.2.2 Asymptotic Normality of the MLE

For nonnegative integers $\alpha_1,\cdots,\alpha_d$, we denote d-index $\alpha = (\alpha_1,\cdots,\alpha_d)$, $|\alpha| = \alpha_1 + \cdots + \alpha_d$ and

$$D^\alpha f(\theta) = \frac{\partial^{\alpha_1}}{\partial\theta^{\alpha_1}}\cdots\frac{\partial^{\alpha_d}}{\partial\theta^{\alpha_d}}f(\theta).$$

In the sequel, we focus on the individual density $\Lambda_1(X^1,\theta)$. To simplify notations, we set $\Lambda(X,\theta) = \Lambda_1(X^1,\theta)$ and $X = X^1$. All possible values of the process X are denoted by x and the integration of $f(X)$ with respect to the measure $\mathbb{P}^1 := d\mu_{X_{\varphi_0,x}}$ is denoted by $\int f(x)dx$. Let $\mathcal{D}^m_{r_0}(\theta_0, d\nu(\varphi))$ denote the class of density functions $g(\varphi,\theta)$ satisfying the following conditions:

C$_1$**:** $g(\varphi,\theta) \in C_b^4(\mathbb{R}^p \times \Theta)$ and $D^\alpha g(\varphi,\theta) \in L^2(d\nu(\varphi))$ for $|\alpha| \leq 4$.

C$_2$**:** There exists $r_0 > 0$ such that for all $\theta \in \overline{B_r(\theta_0)}$ and $|\alpha| \leq 3$, we have

$$\int g(\varphi,\theta_0)^{n+1} g(\varphi,\theta)^{-n} d\nu(\varphi) < \infty \quad \text{and} \quad \int \left(D^\alpha g(\varphi,\theta)\right)^{n+1} g(\varphi,\theta_0)^{-n} d\nu(\varphi) < \infty,$$

for all $n \leq m$ and $r \leq r_0$

Here are some examples of density functions that belong to the class $\mathcal{D}_{r_0}^m(\theta_0, d\varphi)$ where m and r_0 are constants to be specified for each example.

1. Exponential distribution: Let the density function $g(\varphi,\lambda)$ be defined by $g(\varphi,\lambda) = \lambda e^{-\lambda\varphi}$, $\varphi \geq 0$, $\lambda > 0$. Clearly, $g(\varphi,\lambda)$ is infinitely differentiable in λ with integrable derivatives. For every integer $m \geq 1$, the conditions **C**$_1$-**C**$_2$ hold true if $\left(1+\frac{1}{m}\right)^{-1} < \frac{\lambda_0}{\lambda} < 1+\frac{1}{m}$. Hence, we can find an ε-neighborhood of λ_0 on which this condition holds.

2. Gaussian distribution: Let the density function $g(\varphi,\mu,\sigma^2)$ be defined by $g(\varphi,\mu,\sigma^2) = \frac{1}{\sqrt{2\pi\sigma^2}} e^{-\frac{1}{2\sigma^2}(\varphi-\mu)^2}$, $-\infty < \varphi < \infty$, $\mu \in \mathbb{R}$, $\sigma^2 > 0$. For any integer $m \geq 1$, the conditions **C**$_1$-**C**$_2$ hold provided that $\left(1+\frac{1}{m}\right)^{-1} < \frac{\sigma^2}{\sigma_0^2} < 1+\frac{1}{m}$.

3. Cauchy distribution: Let $g(\varphi,\alpha,\beta) = \frac{1}{\pi\beta} \cdot \frac{1}{1+\left(\frac{\varphi-\alpha}{\beta}\right)^2}$, $-\infty < \varphi < \infty$, $\alpha \in \mathbb{R}$, $\beta > 0$. Clearly $g(\varphi,\alpha,\beta) \in \mathcal{D}_{r_0}^m(\alpha_0,\beta_0,d\varphi)$ for all $r_0 > 0$ and $m \geq 1$.

4. Logistic distribution: Let $g(\varphi,\mu) = \frac{e^{-\varphi-\mu}}{(1+e^{-\varphi-\mu})^2}$, $\varphi,\mu \in \mathbb{R}$. $g(\varphi,\mu)$ belongs to the class $\mathcal{D}_{r_0}^m(\mu_0,d\varphi)$ for all $r_0 > 0$ and $m \geq 1$.

5. Gamma distribution: $\Gamma(\varphi,\alpha,\beta)$, $\alpha > 0$, $\beta > 0$, $\varphi \geq 0$ belongs to the class $\mathcal{D}_{r_0}^m(\alpha_0,\beta_0,d\varphi)$ if $r_0 > 0$ and $m \geq 1$ are chosen such that $\left(\frac{\alpha_0}{\alpha}, \frac{\beta}{\beta_0}\right) \in \left(\left(1+\frac{1}{m}\right)^{-1}; 1+\frac{1}{m}\right)^2$.

For the asymptotic normality of the MLE, we make the following assumptions:

A$_7$: There exist $r_0 > 0$ such that $g(\varphi,\theta) \in \mathcal{D}_{r_0}^8(\theta_0, d\nu(\varphi))$

A$_8$: The Fisher information matrix $\mathcal{I}_X(\theta_0)$ is nonsingular.

The following statements hold:

Proposition 4.2.5.

(i) *For* $|\alpha| \leq 3$, $D^\alpha \Lambda(X,\theta) \in L^2(\mathbb{P}_{\theta_0})$ *is continuous in* θ, $\mathbb{P}_{\theta_0}$*-a.s*

(ii) *The Fisher information matrix is finite.*

(iii) *There is some* $r_0 > 0$ *and random function* $H(X,\theta_0)$ *depending only on* θ_0 *such that*

$$\sup_{\theta \in \overline{B_r(\theta_0)}} \left| D^2 \log \Lambda(X,\theta_0) - D^2 \log \Lambda(X,\theta) \right| \leq r H(X,\theta_0), \quad \textit{for } 0 < r < r_0$$

with $\mathbb{E}_{\theta_0} H(X,\theta_0) < \infty$.

Proof. **(i)** Observe that $D^\alpha \Lambda(X,\theta) = \sum_{n\geq 0} J_n(D^\alpha \beta^{\otimes n}(\theta))$ if this equality makes sense, that is, the differentiation can be passed under the integral sign and the series $\sum_{n\geq 0} J_n(D^\alpha \beta^{\otimes n}(\theta))$ is convergent in $L^2(\mathbb{P}_{\theta_0})$. First, we shall show that $J_n\left(D^\alpha \beta^{\otimes n}(\theta)\right) \in L^2(\mathbb{P}_{\theta_0})$ for $|\alpha| \leq 4$. By using BDG inequality, we obtain

$$\begin{aligned}
&\mathbb{E}_{\theta_0}\left\{ \int_{S_n} \int \prod_{l=1}^{n} \beta(X(t_l),\varphi,\varphi_0) D^\alpha g(\varphi,\theta) d\nu(\varphi) dW^{\otimes n}(t) \right\}^2 \\
&\leq C_r^n \int_{S_n} \int \mathbb{E}_{\theta_0}\left\{ \prod_{l=1}^{n} \beta(X(t_l),\varphi,\varphi_0)^2 \right\} (D^\alpha g(\varphi,\theta))^2 d\nu(\varphi) dt^{\otimes n} \\
&\leq 2^{n-1} \frac{(TM^2 C_r)^n}{n!} \left(1 + \mathbb{E}_{\theta_0} \sup_{t\leq T} |X_t|^{2\gamma n}\right) \int (D^\alpha g(\varphi,\theta))^2 d\nu(\varphi) < \infty.
\end{aligned}$$

Thus, the differentiation can be passed under the integral sign and
$D^\alpha J_n\left(\beta^{\otimes n}(\theta)\right) = J_n\left(D^\alpha \beta^{\otimes n}(\theta)\right)$ where

$$D^\alpha \beta^{\otimes n}(\theta)(t_1,\cdots,t_n) = \int \prod_{l=1}^{n} \beta(Xt_l,\varphi,\varphi_0) D^\alpha g(\varphi,\theta) d\nu(\varphi)$$

We have also

$$\begin{aligned}
\left\{ \mathbb{E}_{\theta_0} \left| \sum_{n\geq 0} J_n(D^\alpha \beta^{\otimes n}(\theta)) \right|^2 \right\}^{1/2} &\leq \liminf_{m\longrightarrow\infty} \left\{ \mathbb{E}_{\theta_0} \left| \sum_{n=0}^{m} J_n(D^\alpha \beta^{\otimes n}(\theta)) \right|^2 \right\}^{1/2} \\
&\leq \liminf_{m\longrightarrow\infty} \sum_{n=0}^{m} \left\| J_n(D^\alpha \beta^{\otimes n}(\theta)) \right\|_{L^2(\mathbb{P}_{\theta_0})} \\
&\leq \left(\int (D^\alpha g(\varphi,\theta))^2 d\nu(\varphi) \right)^{1/2} \sum_{n\geq 0} \frac{(2TM^2 C_r)^{n/2}}{\sqrt{n!}} (1 + M_2^{\gamma n}) < \infty.
\end{aligned}$$

Which implies that $\sum_{n\geq 0} J_n(D^\alpha \beta^{\otimes n}(\theta))$ is convergent in $L^2(\mathbb{P}_{\theta_0})$. The aforementioned inequalities are respectively justified by Fatou lemma, Minkowski inequality and the previous result. Since the formula (4.18) holds true for any function $f(\varphi,\theta) = D^\alpha g(\varphi,\theta)$, $|\alpha| \leq 3$, then in similar fashion, we can prove that $D^\alpha \Lambda(X,\theta)$ is continuous in θ, $\mathbb{P}_{\theta_0}$-a.s (see the proof of **(ii)** in Proposition 4.2.2).

(ii) Let $r_0 > 0$, so that $g(\varphi,\theta) \in \mathcal{D}^8_{r_0}(\theta_0, d\nu(\varphi))$. This condition enables us to show that $D^\alpha \log \Lambda(X,\theta) \in L^2(\mathbb{P}_{\theta_0})$ for $1 \leq \alpha \leq 2$ and $\theta \in \Theta$. In particular the Fisher information matrix is finite. In what follows, we will systematically use Hölder's inequality and Fubini theorem. Let $j,k,l \in \{1,\cdots,d\}$ and set $\psi(\varphi) := \psi_k(\varphi,\theta_0) = \frac{\partial g(\varphi,\theta)}{\partial \theta_k}|_{\theta=\theta_0}$. We have

$$\mathbb{E}_{\theta_0}\left(\frac{\partial \log \Lambda(X,\theta)}{\partial \theta_k}|_{\theta=\theta_0}\right)^2 = \mathbb{E}_{\theta_0}\left\{\frac{\left(\int L_T(X,\varphi,\varphi_0)\psi(\varphi)d\nu(\varphi)\right)^2}{\Lambda(X,\theta_0)^2}\right\}$$

$$= \mathbb{E}_{\theta_0}\left\{\frac{\left(\int (L_T(X,\varphi,\varphi_0)g(\varphi,\theta_0))^{1/2}\left[(L_T(X,\varphi,\varphi_0)g(\varphi,\theta_0))^{1/2}(\psi(\varphi)g(\varphi,\theta_0)^{-1})\right]d\nu(\varphi)\right)^2}{\Lambda(X,\theta_0)^2}\right\}$$

$$\leq \mathbb{E}_{\theta_0}\left\{\frac{\int L_T(X,\varphi,\varphi_0)\left(\psi(\varphi)^2 g(\varphi,\theta_0)^{-1}\right)d\nu(\varphi)}{\Lambda(X,\theta_0)}\right\}.$$

Hence,

$$\begin{aligned}
\mathbb{E}_{\theta_0}\left(\frac{\partial \log \Lambda(X,\theta)}{\partial \theta_k}|_{\theta=\theta_0}\right)^2 &\leq \int\int L_T(x,\varphi,\varphi_0)\left(\psi(\varphi)^2 g(\varphi,\theta_0)^{-1}\right)d\nu(\varphi)dx \\
&\leq \int\int L_T(x,\varphi,\varphi_0)\left(\psi(\varphi)^2 g(\varphi,\theta_0)^{-1}\right)dx\,d\nu(\varphi) \\
&\leq \int \left(\psi(\varphi)^2 g(\varphi,\theta_0)^{-1}\right)d\nu(\varphi) < \infty.
\end{aligned}$$

We know that

$$\begin{aligned}
\left(\frac{\partial^2 \log \Lambda(X,\theta)}{\partial \theta_k \partial \theta_l}\right)^2 &\leq 2\frac{\Lambda(X,\theta)^2\left(\frac{\partial^2 \Lambda(X,\theta)}{\partial \theta_k \partial \theta_l}\right)^2 + \left(\frac{\partial \Lambda(X,\theta)}{\partial \theta_k}\right)^2\left(\frac{\partial \Lambda(X,\theta)}{\partial \theta_l}\right)^2}{\Lambda(X,\theta)^4} \\
&\leq 2\frac{\left(\frac{\partial^2 \Lambda(X,\theta)}{\partial \theta_k \partial \theta_l}\right)^2}{\Lambda(X,\theta)^2} + 2\frac{\left(\frac{\partial \Lambda(X,\theta)}{\partial \theta_k}\right)^2\left(\frac{\partial \Lambda(X,\theta)}{\partial \theta_l}\right)^2}{\Lambda(X,\theta)^4}.
\end{aligned} \tag{4.20}$$

We shall prove that the RHS of (4.20) evaluated at θ_0 is of finite expactaion under $\mathbb{P}_{\theta_0}$. We simplify notations by setting $\psi_{k,l}(\varphi,\theta) := \dfrac{\partial^2 g(\varphi,\theta)}{\partial\theta_k\partial\theta_l}$ and $\psi_k(\varphi,\theta) := \dfrac{\partial g(\varphi,\theta)}{\partial\theta_k}$, for $k \in \{1,\cdots,d\}$. Simple computations yield

$$\frac{\left(\frac{\partial^2\Lambda(X,\theta)}{\partial\theta_k\partial\theta_l}\right)^2}{\Lambda(X,\theta)^2} \leq \frac{\int L_T(X,\varphi,\varphi_0)\left(\psi_{k,l}(\varphi,\theta)^2 g(\varphi,\theta_0)^{-1}\right)d\nu(\varphi)}{\Lambda(X,\theta)}.$$

Which implies

$$\mathbb{E}_{\theta_0}\left\{\frac{\left(\frac{\partial^2\Lambda(X,\theta)}{\partial\theta_k\partial\theta_l}\right)^2}{\Lambda(X,\theta)^2}|_{\theta=\theta_0}\right\} \leq \int\left(\psi_{k,l}(\varphi,\theta_0)^2 g(\varphi,\theta_0)^{-1}\right)d\nu(\varphi)\int L_T(x,\varphi,\varphi_0)dx < \infty.$$

Similarly, we have

$$\mathbb{E}_{\theta_0}\left\{\frac{\left(\frac{\partial\Lambda(X,\theta)}{\partial\theta_k}\right)^2\left(\frac{\partial\Lambda(X,\theta)}{\partial\theta_l}\right)^2}{\Lambda(X,\theta)^4}|_{\theta=\theta_0}\right\} \leq$$

$$\leq \int \frac{\left(\int L_T(x,\varphi,\varphi_0)\psi_k(\varphi,\theta_0)d\nu(\varphi)\right)^2}{\Lambda(x,\theta_0)^{3/2}} \times \frac{\left(\int L_T(x,\varphi,\varphi_0)\psi_l(\varphi,\theta_0)d\nu(\varphi)\right)^2}{\Lambda(x,\theta_0)^{3/2}}dx$$

$$\leq \int \frac{\int L_T(x,\varphi,\varphi_0)\left(\psi_k(\varphi,\theta_0)^2 g(\varphi,\theta_0)^{-1}\right)d\nu(\varphi)}{\Lambda(x,\theta_0)^{1/2}} \times \frac{\int L_T(x,\varphi,\varphi_0)\left(\psi_l(\varphi,\theta_0)^2 g(\varphi,\theta_0)^{-1}\right)d\nu(\varphi)}{\Lambda(x,\theta_0)^{1/2}}dx$$

But we know that for $f = k,l$

$$\frac{\int L_T(x,\varphi,\varphi_0)\left(\psi_f(\varphi,\theta_0)^2 g(\varphi,\theta_0)^{-1}\right)d\nu(\varphi)}{\Lambda(x,\theta_0)^{1/2}} \leq \left[\int L_T(x,\varphi,\varphi_0)\left(\psi_f(\varphi,\theta_0)^4 g(\varphi,\theta_0)^{-3}\right)d\nu(\varphi)\right]^{1/2}.$$

Hence,

$$
\mathbb{E}_{\theta_0}\left\{ \frac{\left(\frac{\partial \Lambda(X,\theta)}{\partial \theta_k}\right)^2 \left(\frac{\partial \Lambda(X,\theta)}{\partial \theta_l}\right)^2}{\Lambda(X,\theta)^4} \,|_{\theta=\theta_0} \right\} \leq
$$

$$
\begin{aligned}
&\leq \int \left[\int L_T(x,\varphi,\varphi_0) \left(\psi_k(\varphi,\theta_0)^4 g(\varphi,\theta_0)^{-3}\right) d\nu(\varphi) \right]^{1/2} \\
&\qquad\qquad\qquad \times \left[\int L_T(x,\varphi,\varphi_0) \left(\psi_l(\varphi,\theta_0)^4 g(\varphi,\theta_0)^{-3}\right) d\nu(\varphi) \right]^{1/2} dx \\
&\leq \frac{1}{2} \int\int L_T(x,\varphi,\varphi_0) \left(\psi_k(\varphi,\theta_0)^4 g(\varphi,\theta_0)^{-3}\right) dx d\nu(\varphi) \\
&\qquad\qquad\qquad + \frac{1}{2} \int\int L_T(x,\varphi,\varphi_0) \left(\psi_l(\varphi,\theta_0)^4 g(\varphi,\theta_0)^{-3}\right) dx d\nu(\varphi) \\
&\leq \frac{1}{2} \int \left(\psi_k(\varphi,\theta_0)^4 g(\varphi,\theta_0)^{-3}\right) d\nu(\varphi) + \frac{1}{2} \int \left(\psi_l(\varphi,\theta_0)^4 g(\varphi,\theta_0)^{-3}\right) d\nu(\varphi) < \infty.
\end{aligned}
$$

(iii) Let $\theta \in \overline{B_r(\theta_0)}$. For sufficiently small r, the mean value theorem yields for each $k, j \in \{1, \cdots, d\}$

$$
\sup_{\theta \in \overline{B_r(\theta_0)}} \left| \frac{\partial^2 \log \Lambda(X,\theta_0)}{\partial \theta_k \partial \theta_j} - \frac{\partial^2 \log \Lambda(X,\theta)}{\partial \theta_k \partial \theta_j} \right| \leq r \sum_{l=1}^{d} \left| \frac{\partial^3 \log \Lambda(X,\theta)}{\partial \theta_l \partial \theta_k \partial \theta_j} |_{\theta=\theta^*} \right|,
$$

where θ^* is the maximizer of $\sum_{l=1}^{d} \left| \frac{\partial^3 \log \Lambda(X,\theta)}{\partial \theta_l \partial \theta_k \partial \theta_j} \right|$ which depends only on $\overline{B_r(\theta_0)}$ (that is, it depends only on θ_0). Set $H(X,\theta_0) = \sum_{l=1}^{d} \left| \frac{\partial^3 \log \Lambda(X,\theta)}{\partial \theta_l \partial \theta_k \partial \theta_j} |_{\theta=\theta^*} \right|$. We will prove that $\mathbb{E}_{\theta_0} H(X,\theta_0) < \infty$. First, note that

$$
\begin{aligned}
\frac{\partial^3 \log \Lambda(X,\theta)}{\partial \theta_l \partial \theta_k \partial \theta_j} &= \sum_{i=1}^{5} G_i(X,\theta), \\
\text{where} \quad G_1(X,\theta) &= \Lambda(X,\theta)^{-1} \frac{\partial^3 \Lambda(X,\theta)}{\partial \theta_l \partial \theta_k \partial \theta_j} \\
G_2(X,\theta) &= -\Lambda(X,\theta)^{-2} \frac{\partial^2 \Lambda(X,\theta)}{\partial \theta_l \partial \theta_k} \frac{\partial \Lambda(X,\theta)}{\partial \theta_j} \\
G_3(X,\theta) &= -\Lambda(X,\theta)^{-2} \frac{\partial^2 \Lambda(X,\theta)}{\partial \theta_l \partial \theta_j} \frac{\partial \Lambda(X,\theta)}{\partial \theta_k} \\
G_4(X,\theta) &= -\Lambda(X,\theta)^{-2} \frac{\partial^2 \Lambda(X,\theta)}{\partial \theta_k \partial \theta_j} \frac{\partial \Lambda(X,\theta)}{\partial \theta_l} \\
G_5(X,\theta) &= 2\Lambda(X,\theta)^{-3} \frac{\partial \Lambda(X,\theta)}{\partial \theta_l} \frac{\partial \Lambda(X,\theta)}{\partial \theta_k} \frac{\partial \Lambda(X,\theta)}{\partial \theta_j}.
\end{aligned}
$$

We are going to show that $\mathbb{E}_{\theta_0}|G_1(X,\theta)|$, $\mathbb{E}_{\theta_0}|G_2(X,\theta)|$ and $\mathbb{E}_{\theta_0}|G_5(X,\theta)|$ are finite for all $\theta \in \overline{B_r(\theta_0)}$ and $0 < r < r_0$. To simplify notations, we set $\psi(\varphi,\theta) := \psi_{l,k,j}(\varphi,\theta) = \dfrac{\partial^3 g(\varphi,\theta)}{\partial\theta_l\partial\theta_k\partial\theta_j}$. By using Hölder's inequality, we obtain

$$G_1(X,\theta)^2 = \frac{\left(\int L_T(X,\varphi,\varphi_0)\psi(\varphi,\theta)d\nu(\varphi)\right)^2}{\Lambda(X,\theta)}$$
$$\leq \frac{\left(\int [L_T(X,\varphi,\varphi_0)g(\varphi,\theta_0)]^{3/4}[(L_T(X,\varphi,\varphi_0)g(\varphi,\theta_0))^{1/4}\psi(\varphi,\theta)g(\varphi,\theta_0)^{-1}]d\nu(\varphi)\right)^2}{\Lambda(X,\theta)^2}$$
$$\leq \frac{\left(\int L_T(X,\varphi,\varphi_0)g(\varphi,\theta_0)d\nu(\varphi)\right)^{3/2}\left(\int L_T(X,\varphi,\varphi_0)\psi(\varphi,\theta)^4 g(\varphi,\theta_0)^{-3}d\nu(\varphi)\right)^{1/2}}{\Lambda(X,\theta)^2}.$$

Thus

$$\mathbb{E}_{\theta_0}G_1(X,\theta)^2 \leq \int \frac{\Lambda(x,\theta_0)^{5/2}\left[\int L_T(x,\varphi,\varphi_0)\psi(\varphi,\theta)^4 g(\varphi,\theta_0)^{-3}d\nu(\varphi)\right]^{1/2}}{\Lambda(x,\theta)^2}dx$$
$$\leq \frac{1}{2}\int \frac{\Lambda(x,\theta_0)^5}{\Lambda(x,\theta)^4}dx + \frac{1}{2}\int\int L_T(x,\varphi,\varphi_0)\psi(\varphi,\theta)^4 g(\varphi,\theta_0)^{-3}d\nu(\varphi)dx$$
$$\leq \frac{1}{2}\int g(\varphi,\theta_0)^5 g(\varphi,\theta_0)^{-4}d\nu(\varphi) + \frac{1}{2}\int \psi(\varphi,\theta)^4 g(\varphi,\theta_0)^{-3}d\nu(\varphi).$$

The first term on the RHS of the last inequality is obtained by using Hölder's inequality as follows

$$\frac{\Lambda(x,\theta_0)^5}{\Lambda(x,\theta)^4} = \frac{\left(\int L_T(x,\varphi,\varphi_0)g(\varphi,\theta_0)d\nu(\varphi)\right)^5}{\Lambda(x,\theta)^4}$$
$$\leq \frac{\left[\int L_T(x,\varphi,\varphi_0)g(\varphi,\theta_0)d\nu(\varphi)\right]^4 \int L_T(x,\varphi,\varphi_0)g(\varphi,\theta_0)^5 g(\varphi,\theta_0)^{-4}d\nu(\varphi)}{\Lambda(x,\theta)^4}$$
$$\leq \int L_T(x,\varphi,\varphi_0)g(\varphi,\theta_0)^5 g(\varphi,\theta_0)^{-4}d\nu(\varphi).$$

Hence, $\mathbb{E}_{\theta_0}G_1(X,\theta)^2 < \infty$ for all $\theta \in \overline{B_r(\theta_0)}$ and $0 < r < r_0$. Similarly we have

$$\mathbb{E}_{\theta_0}|G_2(X,\theta)| \leq \mathbb{E}_{\theta_0}\left\{\left|\Lambda(X,\theta)^{-1}\frac{\partial^2\Lambda(X,\theta)}{\partial\theta_l\partial\theta_k}\right|\left|\Lambda(X,\theta)^{-1}\frac{\partial\Lambda(X,\theta)}{\partial\theta_j}\right|\right\}$$
$$\leq \frac{1}{2}\mathbb{E}_{\theta_0}\left(\Lambda(X,\theta)^{-1}\frac{\partial^2\Lambda(X,\theta)}{\partial\theta_l\partial\theta_k}\right)^2 + \frac{1}{2}\mathbb{E}_{\theta_0}\left(\Lambda(X,\theta)^{-1}\frac{\partial\Lambda(X,\theta)}{\partial\theta_j}\right)^2$$
$$\leq \frac{1}{2}\int g(\varphi,\theta_0)^5 g(\varphi,\theta_0)^{-4}d\nu(\varphi) + \frac{1}{2}\int\left[\psi_{l,k}(\varphi,\theta)^4 + \psi_j(\varphi,\theta)^4\right]g(\varphi,\theta_0)^{-3}d\nu(\varphi) < \infty,$$

where $\dfrac{\partial^2 g(\varphi,\theta)}{\partial\theta_l\partial\theta_k} = \psi_{l,k}(\varphi,\theta)$ and $\dfrac{\partial g(\varphi,\theta)}{\partial\theta_j} = \psi_j(\varphi,\theta)$. By using the same techniques, we prove that $G_3(X,\theta)$ and $G_4(X,\theta)$ are of finite expectation for all $\theta \in \overline{B_r(\theta_0)}$ and $0 < r < r_0$.

Set $\dfrac{\partial g(\varphi,\theta)}{\partial\theta_l} = \psi_l(\varphi,\theta)$, $\dfrac{\partial g(\varphi,\theta)}{\partial\theta_k} = \psi_k(\varphi,\theta)$ and $\dfrac{\partial g(\varphi,\theta)}{\partial\theta_j} = \psi_j(\varphi,\theta)$. By using the fact that $2ab \le a^2 + b^2,\ a,b \in \mathbb{R}$, we obtain

$$\mathbb{E}_{\theta_0}\left|G_5(X,\theta)\right| \le \mathbb{E}_{\theta_0}\left(\Lambda(X,\theta)^{-1}\frac{\partial\Lambda(X,\theta)}{\partial\theta_l}\right)^2 + \mathbb{E}_{\theta_0}\left(\Lambda(X,\theta)^{-2}\frac{\partial\Lambda(X,\theta)}{\partial\theta_k}\frac{\partial\Lambda(X,\theta)}{\partial\theta_j}\right)^2$$

Then, with the same techniques used previously, we state

$$\begin{aligned}\mathbb{E}_{\theta_0}\left|G_5(X,\theta)\right| \;&\le\; \frac{1}{2}\int g(\varphi,\theta_0)^5 g(\varphi,\theta)^{-4}d\nu(\varphi) + \frac{1}{2}\int \psi_l(\varphi,\theta)^4 g(\varphi,\theta_0)^{-3}d\nu(\varphi)\\ &\quad+\frac{1}{2}\mathbb{E}_{\theta_0}\left(\frac{\left(\frac{\partial\Lambda(X,\theta)}{\partial\theta_k}\right)^4}{\Lambda(X,\theta)^4}\right) + \frac{1}{2}\mathbb{E}_{\theta_0}\left(\frac{\left(\frac{\partial\Lambda(X,\theta)}{\partial\theta_j}\right)^4}{\Lambda(X,\theta)^4}\right).\end{aligned} \tag{4.21}$$

It remains to show that the last two terms on the RHS of (4.21) are finite. Again, by using Holder's inequality, we obtain

$$\begin{aligned}&\frac{\left(\frac{\partial\Lambda(X,\theta)}{\partial\theta_k}\right)^4}{\Lambda(X,\theta)^4} = \frac{\left(\int L_T(X,\varphi,\varphi_0)\psi_k(\varphi,\theta)d\nu(\varphi)\right)^4}{\Lambda(X,\theta)^4}\\ &\le \frac{\left(\int L_T(X,\varphi,\varphi_0)g(\varphi,\theta_0)d\nu(\varphi)\right)^{7/2}\left(\int L_T(X,\varphi,\varphi_0)\psi_k(\varphi,\theta)^8 g(\varphi,\theta_0)^{-7}d\nu(\varphi)\right)^{1/2}}{\Lambda(X,\theta)^4}.\end{aligned}$$

Therefore

$$\begin{aligned}\mathbb{E}_{\theta_0}\left(\frac{\left(\frac{\partial\Lambda(X,\theta)}{\partial\theta_k}\right)^4}{\Lambda(X,\theta)^4}\right) \;&\le\; \int \frac{\Lambda(x,\theta_0)^{9/2}\left[\int L_T(x,\varphi,\varphi_0)\psi_k(\varphi,\theta)^8 g(\varphi,\theta_0)^{-7}d\nu(\varphi)\right]^{1/2}}{\Lambda(x,\theta)^4}dx\\ &\le\; \frac{1}{2}\int\frac{\Lambda(x,\theta_0)^9}{\Lambda(x,\theta)^8}dx + \frac{1}{2}\int\int L_T(x,\varphi,\varphi_0)\psi_k(\varphi,\theta)^8 g(\varphi,\theta_0)^{-7}dx\,d\nu(\varphi).\end{aligned} \tag{4.22}$$

Since

$$\begin{aligned}\Lambda(x,\theta_0)^9 \;&=\; \left(\int L_T(x,\varphi,\varphi_0)g(\varphi,\theta_0)d\nu(\varphi)\right)^9\\ &\le\; \Lambda(x,\theta)^8\int L_T(x,\varphi,\varphi_0)g(\varphi,\theta_0)^9 g(\varphi,\theta)^{-8}d\nu(\varphi),\end{aligned}$$

then

$$\int \frac{\Lambda(x,\theta_0)^9}{\Lambda(x,\theta)^8}dx \leq \int\int L_T(x,\varphi,\varphi_0)g(\varphi,\theta_0)^9 g(\varphi,\theta)^{-8}dxd\nu(\varphi) \leq \int g(\varphi,\theta_0)^9 g(\varphi,\theta)^{-8}d\nu(\varphi).$$

Hence, from (4.22) it follows that

$$\mathbb{E}_{\theta_0}\left(\frac{\left(\frac{\partial\Lambda(X,\theta)}{\partial\theta_k}\right)^4}{\Lambda(X,\theta)^4}\right) \leq \frac{1}{2}\int g(\varphi,\theta_0)^9 g(\varphi,\theta)^{-8}d\nu(\varphi) + \frac{1}{2}\int \psi_k(\varphi,\theta)^8 g(\varphi,\theta_0)^{-7}d\nu(\varphi) < \infty.$$

Similarly, if we replace k by j we obtain

$$\mathbb{E}_{\theta_0}\left(\frac{\left(\frac{\partial\Lambda(X,\theta)}{\partial\theta_j}\right)^4}{\Lambda(X,\theta)^4}\right) < \infty,$$

which completes the proof of Proposition 4.2.5. □

At this stage, all conditions needed in Theorem 3.0.6 to establish the asymptotic normality of the MLE are fulfilled (Proposition 4.2.5, Theorem 4.2.3 and the hypothesis $\mathbf{A}_9$). Thus the following result holds.

Theorem 4.2.6. *Under the assumptions $\boldsymbol{A}_1$-$\boldsymbol{A}_8$, the MLE $\widehat{\theta}_N$ is asymptotically normal. i.e.,*

$$\sqrt{N}\left(\widehat{\theta}_N - \theta_0\right) \xrightarrow{\mathcal{L}} \mathcal{N}\left(0, \mathcal{I}_{X^1}(\theta_0)^{-1}\right), \quad \textit{under } \mathbb{P}_{\theta_0}$$

As mentioned before, the most delicate task in proving the main previous results was to show that $\Lambda(X,\theta)$ is continous in θ, $\mathbb{P}_{\theta_0}$-a.s; the difficulty relies on finding an upper-bound of $\mathbb{E}_{\theta_0}\prod_{l=1}^{n}\beta(X(t_l),\varphi,\varphi_0)^{2r}$ (see the inequality (4.19)). However, it is possible to prove that $\Lambda(X,\theta)$ is continous in θ, $\mathbb{P}_{\theta_0}$-a.s under weakned assumptions (see Appendix B. (p.111)), in the sense that $\mathbf{A}_4$ and $\mathbf{A}_5$ are replaced by the following assumptions

A'$_4$: There exist a nonnegative constants M (which may depend on φ_0), γ_1 and γ_2 such that

$$|\beta(x,\varphi,\varphi_0)| \leq M\left(1 + |x|^{\gamma_1} + \|\varphi\|^{\gamma_2}\right)$$

and for all $c > 0$, $\sum_{n\geq 0}\frac{c^n}{\sqrt[2r]{n!}}\mu_{n,\gamma_2,r}^{1/2r} < \infty$, where $\mu_{n,\gamma_2,r} = \int \|\varphi\|^{2\gamma_2 nr}h(\varphi)d\nu(\varphi)$ with $2r\delta - d > 0$ and $h(\varphi)$ is the density function appearing in $\mathbf{A}_6$.

A'$_5$ There exists $M_3 > 0$ such that $\sup_{t\leq T}\mathbb{E}_{\theta_0}\left|X^i(t)\right|^{2k} \leq M_3^k$, for all $k > 0$ and $i = 1, \cdots, N$.

CHAPTER 5
STATISTICAL INFERENCE FOR FRACTIONAL STOCHASTIC PROCESSES WITH RANDOM EFFECTS

In this chapter, we try to extende the results given in Chapter 4 to fractional diffusion processes with random effects. To do this, we begin with fractional stochastic differential equations (FSDE) with general linear drift containing random effects. More precisely, we focuss on two kind of FSDE's

- FSDE's with additive random effects in the drift.
- FSDE's with multiplicative random effects in the drift and small fractional diffusion.

5.1 PROBLEM OUTLINE

Consider the following FSDE's

$$\begin{aligned} dX^i(t) &= \big(a(X^i(t)) + \phi_i b(t, X^i(t))\big)\, dt + \sigma(t, X^i(t)) dW^{H,i}(t), \quad t \leq T, \\ X^i(0) &= x_i \in \mathbb{R}, \quad i = 1, \cdots, N, \end{aligned} \tag{5.1}$$

where $\phi_1, \cdots, \phi_N$ are i.i.d $\mathbb{R}$-valued random variables (random effects) with comon density f, $\big(W^{H,1}, \cdots, W^{H,N}\big)$ are independent standard fractional Brownian motions with common Hurst parameter $H \in (0,1)$ and $\phi_1, \cdots, \phi_N$ are independent of $\big(W^{H,1}, \cdots, W^{H,N}\big)$. First, we consider the case (Section 5.2) where $b(t,x) = \sigma(t,x) = 1$ and $f \equiv \mathcal{N}(\mu, \omega^2)$. The estimators $\widehat{H}$, $\widehat{\mu}$ and $\widehat{\omega^2}$ of H, μ and ω^2 are constructed and examined. The asymptotic properties are studied as the number of subjects tends to infinity. Our results are illustrated by numerical examples. Second, when f is non-parametric, we provide a class of estimates (Sections 5.3 - 5.4) and study their L^p-risk ($p = 1,$ or 2) and/or pointwise-risk for both cases

- **Case 1 :** $b(t,x) = b(t)$ and $\sigma(t,x) = \sigma(t)$ with $N, T \to \infty$
- **Case 2 :** $a(x) = 0$, $b(t,x) = b(x)$ and $\sigma(t,x) = \varepsilon$ with $N \to \infty$ and $\varepsilon \to 0$.

5.2 PARAMETRIC ESTIMATION OF POPULATION PARAMETERS IN FSDE'S WITH ADDITIVE RANDOM EFFECTS

This section deals with the problem of inference associated with linear fractional diffusion process with random effects in the drift. In particular we are concerned with the maximum likelihood estimators (MLE) of the random effect parameters. First of all, we estimate the Hurst parameter $H \in (0,1)$ from one single subject. Second, assuming the Hurst index $H \in (0,1)$ is known, we derive the MLE's and examine their asymptotic behaviour as the number of subjects under study becomes large, with random effects normally distributed.

5.2.1 Model, Notations and Preliminary Results

Consider the N subjects $\big(X^i(t), \mathcal{F}_t^i, t \le T\big)$ with dynamics ruled by (5.1) with $b(t,x) = \sigma(t,x) = 1$. The functions $a(\cdot)$ and $b(\cdot)$ are supposed to be known in their own spaces. Let the random effects ϕ_i be $\mathcal{F}_0^i$-measurable with common density $f(\varphi,\theta)d\nu(\varphi)$, where ν is some dominating measure on $\mathbb{R}$ and θ is unknown parameter. Set $\theta \in U$, where U is an open set in $\mathbb{R}^d$. Sufficient conditions for the existence and uniqueness of solutions to (5.1) are given in Theorem 2.2.3, and more details can be found in [80, p. 197], [86], and references therein.

Let C_T denote the space of real continuous functions $(x(t) \ : \ t \in [0,T])$ defined on $[0,T]$ endowed with σ-field $\mathcal{B}_T$. The σ-field $\mathcal{B}_T$ is associated with the topology of uniform convergence on $[0,T]$. We introduce the distribution $\mu_{X^i_{\varphi,H}}$ on $(C_T, \mathcal{B}_T)$ of the process $\big(X^i|\phi_i = \varphi\big)$. On $\mathbb{R}\times C_T$, $Q^i_{\theta,H} = f(\varphi,\theta)d\nu \otimes \mu_{X^i_{\varphi,H}}$ denote the joint distribution of (ϕ_i, X^i). Let $\mathbb{P}^i_{\theta,H}$ be the marginal distribution of $\big(X^i(t) \ : \ t \le T\big)$ on $(C_T, \mathcal{B}_T)$. Since the subjects are independent (this is inherited from the independence of ϕ_i and $W^{H,i}$), the distribution of the whole sample $\big(X^i(t) \ : \ t \le T, \ i = 1, \cdots, N\big)$ on $C_T^{\otimes N}$ is defined by $\mathbb{P}_{\theta,H} = \otimes_{i=1}^N \mathbb{P}^i_{\theta,H}$. Thus the likelihood can be defined as

$$\Lambda(\theta, H) = \frac{\mathrm{d}\mathbb{P}_{\theta,H}}{\mathrm{d}\mathbb{P}} = \prod_{i=1}^N \frac{d\mathbb{P}^i_{\theta,H}}{d\mathbb{P}^i},$$

where $\mathbb{P} = \otimes_{i=1}^N \mathbb{P}^i$ and $\mathbb{P}^i = \mu_{X^i_{\varphi_0,H}}$, provided that $\mu_{X^i_{\varphi,H}} \ll \mu_{X^i_{\varphi_0,H}}$ for some fixed $\varphi_0 \in \mathbb{R}$. It is well known that $\mu_{X^i_{\varphi,H}}$ coincides with the distribution of the process $X^{i,\varphi}$ defined by:

$$dX^{i,\varphi}(t) = \big(a(X^i(t)) + \varphi b(X^{i,\varphi}(t))\big)\, dt + dW^{H,i}(t), \ \ X^{i,\varphi}(0) = x^i,$$

when $H = 1/2$, since in this case the process (X^i, ϕ_i) is markovian (e.g., [46]); hence, the Girshanov formula can be applied to get the derivative $\dfrac{d\mu_{X^i_{\varphi,H}}}{d\mu_{X^i_{\varphi_0,H}}}$. When $H \ne 1/2$, the non Markovian property of the coupled process (X^i, ϕ_i) makes the construction of the likelihood very difficult. But in our case, the process X^i is transformed into a Y^i for which the law of

$\left(Y^i|\phi_i = \varphi\right)$ coincides with the distribution of a φ-parametrized fractional diffusion process $Y^{i,\varphi}$.

5.2.2 Construction of Estimators and their Asymptotic Properties

Consider the following process

$$\begin{aligned} Y^i(t) &:= X^i(t) - x^i - \int_0^t a(X^i(s))ds, \quad t \geq 0 &(5.2)\\ &= t\phi_i + W^{H,i}(t) \sim \mathcal{N}\left(t\mu, t^2\omega^2 + t^{2H}\right), \quad t \geq 0. &(5.3)\end{aligned}$$

Since ϕ_i and $W^{H,i}$ are independent $\left(Y^i(t) \;:\; t \in [0,T]\right)$ is a Gaussian process. Furthermore, for each $\varphi \in \mathbb{R}$, we have $\mathbb{E}\left(Y^i(t)|\phi_i = \varphi\right) = t\varphi$ and $\mathbb{C}ov\left(Y^i(t), Y^i(s)|\phi_i = \varphi\right) = \frac{1}{2}(t^{2H} + s^{2H} - |t-s|^{2H})$. For each subject Y^i, we consider n observations $Y^i := \left(Y^i(t_1), \cdots, Y^i(t_n)\right)'$ where $0 = t_0 < t_1 < \cdots < t_n = T$ is a subdivision of $[0,T]$. The density of Y^i given $\phi_i = \varphi$ is expressed as

$$\Pi(Y^i|\phi_i = \varphi, H) = \frac{1}{\sqrt{(2\pi)^n \mathrm{det}V(H)}} e^{-\frac{1}{2}(Y^i - \varphi u)' V^{-1}(H)(Y^i - \varphi u)},$$

where $u = (t_1, \cdots, t_n)'$ and $(V(H))_{k,l} = \mathbb{C}ov\left(Y^i(t_k), Y^i(t_l)|\phi_i = \varphi\right)$ is the common covariance matrix of the subjects Y^i, $i = 1, \cdots, N$. The log-likelihood of the whole sample $\left(Y^1, \cdots, Y^N\right)$ is defined as

$$l(\theta, H) = \sum_{i=1}^N \log \int \Pi(Y^i|\phi_i = \varphi, H) f(\varphi, \theta) d\nu(\varphi). \tag{5.4}$$

For a specific distribution (say $f(\varphi,\theta)d\nu(\varphi) = \mathcal{N}(\mu, \omega^2)$), we can solve the integrals given in (5.4). Indeed,

$$\begin{aligned}\int \Pi(Y^i|\phi_i = \varphi, H) f(\varphi, \theta) d\nu(\varphi) &= (2\pi)^{-n/2}\omega^{-1}\mathrm{det}(V(H))^{-1/2}\left(u'V^{-1}(H)u + 1/\omega^2\right)^{-1/2}\\ &\times \exp\left[-\frac{1}{2}\left(\mu^2/\omega^2 + Y^{i\prime}V^{-1}(H)Y^i - \frac{(u'V^{-1}(H)Y^i + \mu/\omega^2)^2}{u'V^{-1}(H)u + 1/\omega^2}\right)\right]. \end{aligned} \tag{5.5}$$

5.2.3 Estimation of the Hurst Parameter H

Using data induced by one single subject (without loss of generality, say Y^1 with $t_j = \frac{j}{n}, j = 1, \cdots, n,\ T = 1$), we may construct a class of estimators of the Hurst index H. More precisely, for all $k > 0$ and for any filter $\gamma = (\gamma_0, \cdots \gamma_l)$ of order $p \geq 2$, that is,

$$\text{for all indices } 0 \leq r < p;\ \sum_{j=0}^l j^r \gamma_j = 0 \ \text{ and } \ \sum_{j=0}^l j^p \gamma_j \neq 0, \tag{5.6}$$

we set

$$
\begin{aligned}
\widehat{H}(n,p,k,\gamma,Y^1) &= g^{-1}_{n,k,\gamma}\left(S_n(k,\gamma)\right), \\
\text{where } S_n(k,\gamma) &= \frac{1}{n-l}\sum_{i=l}^{n-1}\left|\sum_{q=0}^{l}\gamma_q Y^1\left(\frac{i-q}{n}\right)\right|^k, \; g_{n,k,\gamma}(t) = \frac{1}{n^{tk}}\{\pi_t^\gamma(0)\}^{k/2}E_k, \text{ and} \\
\pi_t^\gamma(j) &= -\frac{1}{2}\sum_{q,r}^{l}\gamma_q\gamma_r\left|q-r+j\right|^{2t}, \text{ with } E_k = 2^{k/2}\Gamma(k+1/2)/\Gamma(1/2) \\
\text{and} \quad & \Gamma(x) \text{ is the usual gamma function.}
\end{aligned}
$$

For invertibility of the function $g_{n,k,\gamma}(\cdot)$, we refer to [19, p. 7].

Theorem 5.2.1. *As the number of observations $n \longrightarrow \infty$, the following statements holds true,*

(i) $\widehat{H}(n,p,k,\gamma,Y^1) \overset{\mathbb{P}-as}{\Longrightarrow} H$

(ii) $n^{-1/2}\log(n)\left(\widehat{H}(n,p,k,\gamma,Y^1) - H\right) \overset{\mathcal{L}}{\Longrightarrow} \mathcal{N}\left(0, \dfrac{A(H,k,\gamma)}{k^2}\right)$, *where*

$$
\begin{aligned}
A(t,k,\gamma) &= \sum_{j\geq 1}(c_{2j}^k)^2(2j)!\sum_{i\in\mathbb{Z}}\rho_t^\gamma(i)^{2j}, \; \textit{with} \\
c_{2j}^k &= \frac{1}{(2j)!}\prod_{q=0}^{j-1}(k-2q), \; \textit{and } \rho_t^\gamma(i) = \frac{\pi_t^\gamma(i)}{\pi_t^\gamma(0)}.
\end{aligned}
$$

Proof. Following Coeurjolly [19], we set $V^\gamma(i/n) = \sum_{q=0}^{l}\gamma_q W^{H,1}\left(\frac{i-q}{n}\right)$, for $i = l, \cdots, n-1$. Since the filter γ is of order $p \geq 2$ (see, (5.6)), we have $\sum_{q=0}^{l}\frac{i-q}{n}\gamma_q = 0$. Then, substituting $Y^1\left(\frac{i-q}{n}\right)$ by $\frac{i-q}{n}\phi_1 + W^{H,1}\left(\frac{i-q}{n}\right)$, we obtain

$$
\begin{aligned}
S_n(k,\gamma) &= \frac{1}{n-l}\sum_{i=l}^{n-1}\left|\sum_{q=0}^{l}\gamma_q Y^1\left(\frac{i-q}{n}\right)\right|^k \\
&= \frac{1}{n-l}\sum_{i=l}^{n-1}\left|\sum_{q=0}^{l}\gamma_q\frac{i-q}{n}\phi_1 + \sum_{q=0}^{l}\gamma_q W^{H,1}\left(\frac{i-q}{n}\right)\right|^k \\
&= \frac{1}{n-l}\sum_{i=l}^{n-1}\left|V^\gamma(i/n)\right|^k.
\end{aligned}
$$

Hence, our estimators coincide with estimators $\widehat{H}$ based on k-variations of the fBm (see, [19, Proposition 2]) and the proof is complete. □

5.2.4 Estimation of the Population Parameter $\theta = (\mu, \omega^2)$

Now, assume that H is known. From the log-likelihood given by (5.4) and (5.5), we derive an estimator $\widehat{\mu}$ defined by

$$\widehat{\mu} = \frac{\frac{1}{N}\sum_{i=1}^{N} u'V^{-1}(H)Y^i}{u'V^{-1}(H)u}. \tag{5.7}$$

For the parameter ω^2 it sounds very difficult to derive an estimator. However, we can construct an alternative estimator and study its asymptotic behaviour. Observing that $\widehat{\mu}$ is a sample mean drawn from a sequence of i.i.d random variables, one might think that sample variance could also be used to estimate ω^2. Unfortunately, simple computations shows that such a sample variance is not consistent. Thus, as an alternative we propose the following estimator for ω^2:

$$\widehat{\omega^2} = \frac{1}{N}\sum_{i=1}^{N}\left(\frac{u'V^{-1}(H)Y^i}{u'V^{-1}(H)u}\right)^2 - \frac{1}{N^2}\left(\sum_{i=1}^{N}\frac{u'V^{-1}(H)Y^i}{u'V^{-1}(H)u}\right)^2 - \left(u'V^{-1}(H)u\right)^{-1}. \tag{5.8}$$

Theorem 5.2.2. *The estimator $\widehat{\mu}$ is unbaised, $\widehat{\mu} \overset{\mathbb{P}-as}{\Longrightarrow} \mu$ and $\mathbb{V}ar(\widehat{\mu}) \longrightarrow 0$ as $N \to \infty$.*

Proof. Set $\epsilon^i = \left(W^{H,i}(t_1), \cdots, W^{H,i}(t_n)\right)'$. Substituting Y^i by $\phi_i u + \epsilon^i$, we have $\widehat{\mu} = \frac{1}{N}\sum_{i=1}^{N}\phi_i + \frac{\frac{1}{N}\sum_{i=1}^{N} u'V^{-1}(H)\epsilon^i}{u'V^{-1}(H)u}$, so, $\mathbb{E}(\widehat{\mu}) = \mu$. For the second statement, we consider the random variables $\xi_i(H)$ defined by

$$\xi_i(H) = \frac{u'V^{-1}(H)Y^i}{u'V^{-1}(H)u}. \tag{5.9}$$

Clearly, $\xi_i(H)$ are i.i.d random variables with $\mathbb{E}(\xi_i(H)) = \mu < \infty$, then by strong law of large numbers (e.g., [101, Corollary 1.63]), $\widehat{\mu}$ converges almost surely to μ as $N \to \infty$. Set

$z(H) := (z_1(H), \cdots, z_n(H)) = u'V^{-1}(H)$, we have

$$
\begin{aligned}
\mathbb{V}ar(\widehat{\mu}) &= \mathbb{V}ar\left(\frac{1}{N}\sum_{i=1}^{N}\phi_i\right) + \frac{1}{N^2(z(H)\cdot u)^2}\mathbb{V}ar\left(\sum_{i=1}^{N} z(H)\cdot\epsilon^i\right) \\
&= \frac{1}{N^2}\mathbb{V}ar\left(\sum_{i=1}^{N}\phi_i\right) + \frac{1}{N^2(z(H)\cdot u)^2}\sum_{i,j}^{N}\mathbb{E}\left\{\left(\sum_{k=1}^{n} z_k(H)W^{H,i}(t_k)\right)\left(\sum_{l=1}^{n} z_l(H)W^{H,j}(t_l)\right)\right\} \\
&= \frac{1}{N^2}\sum_{i=1}^{N}\mathbb{V}ar(\phi_i) + \frac{1}{N^2(z(H)\cdot u)^2}\sum_{i,j}^{N}\sum_{k,l}^{n} z_k(H)z_l(H)\mathbb{E}\left(W^{H,i}(t_k)W^{H,j}(t_l)\right) \\
&= \frac{\omega^2}{N} + \frac{1}{N^2(z(H)\cdot u)^2}\sum_{i}^{N}\sum_{k,l}^{n} z_k(H)z_l(H)\mathbb{E}\left(W^{H,i}(t_k)W^{H,i}(t_l)\right) \\
&= \frac{\omega^2}{N} + \frac{1}{N^2(z(H)\cdot u)^2}\sum_{i}^{N}\sum_{k,l}^{n}\frac{1}{2}z_k(H)z_l(H)\left(t_k^{2H} + t_l^{2H} - |t_k - t_l|^{2H}\right) \\
&= \frac{\omega^2}{N} + \frac{1}{N^2(z(H)\cdot u)^2}Nz(H)V(H)z(H)' = \omega^2 + \frac{Nu'V^{-1}(H)V(H)V^{-1}(H)u}{N^2(z(H)\cdot u)^2} \\
&= \frac{\omega^2}{N} + \frac{1}{Nu'V^{-1}(H)u} \to 0 \quad \text{as} \quad N \to \infty.
\end{aligned}
$$

□

Before, we establish the bias of $\widehat{\omega^2}$ the estimator of ω^2, we first give the following result:

Lemma 5.2.3.

$$
\mathbb{E}(\xi_1(H))^2 = \omega^2 + \mu^2 + \frac{1}{u'V^{-1}(H)u} \quad \textit{and} \quad \mathbb{E}\left(\sum_{i=1}^{N}\xi_i(H)\right)^2 = N\omega^2 + N^2\mu^2 + \frac{N}{u'V^{-1}(H)u},
$$

where $\xi_i(H)$ are random variables defined in (5.9).

Proof. Substituting Y^1 by $\phi_1 u + \epsilon^1$ and using the independence of ϕ_1 and ϵ^1, we have

$$
\begin{aligned}
\mathbb{E}(\xi_1(H))^2 &= \mathbb{E}\left(\phi_1 + \frac{u'V^{-1}(H)\epsilon^1}{u'V^{-1}(H)u}\right)^2 \\
&= \mathbb{E}\phi_1^2 + \mathbb{E}\left(\frac{u'V^{-1}(H)\epsilon^1}{u'V^{-1}(H)u}\right)^2 \\
&= \omega^2 + \mu^2 + \frac{1}{u'V^{-1}(H)u}.
\end{aligned}
$$

For the last equality we used the same techniques as in the proof of Theorem 5.2.2. For the second statement; by using the random variables $z_i(H)'$s defined previously, we have

$$\begin{aligned}
\mathbb{E}\left(\sum_{i=1}^{N}\xi_i(H)\right)^2 &= \mathbb{E}\left(\sum_{i=1}^{N}\phi_i+\sum_{i=1}^{N}\frac{z(H)\cdot\epsilon^i}{z(H)\cdot u}\right)^2\\
&= \mathbb{E}\left(\sum_{i=1}^{N}\phi_i\right)^2+\mathbb{E}\left(\sum_{i=1}^{N}\frac{z(H)\cdot\epsilon^i}{z(H)\cdot u}\right)^2\\
&= \sum_{i=1}^{N}\mathbb{E}\phi_i^2+2\sum_{i<j}^{N}\mathbb{E}(\phi_i\phi_j)+\frac{1}{(u'V^{-1}(H)u)^2}\mathbb{V}ar\left(\sum_{i=1}^{N}z(H)\cdot\epsilon^i\right)\\
&= N\omega^2+N^2\mu^2+\frac{N}{u'V^{-1}(H)u}.
\end{aligned}$$

□

Theorem 5.2.4. *The estimator* $\widehat{\omega^2}$ *is asymptotically unbiased,* $\widehat{\omega^2}\overset{\mathbb{P}-as}{\Longrightarrow}\omega^2$ *and* $\mathbb{V}ar(\widehat{\omega^2})=\frac{2(N-1)}{N^2}\left(\omega^2+\frac{1}{u'V^{-1}(H)u}\right)^2\longrightarrow 0,$ *as* $N\to\infty$*.*

Proof. By vitrue of Lemma 5.2.3, we have

$$\begin{aligned}
\mathbb{E}(\widehat{\omega^2}) &= \frac{1}{N}\sum_{i=1}^{N}\left(\omega^2+\mu^2+\frac{1}{u'V^{-1}(H)u}\right)-\frac{1}{N^2}\left(N\omega^2+N^2\mu^2+\frac{N}{u'V^{-1}(H)u}\right)-\frac{1}{u'V^{-1}(H)u}\\
&= \frac{N-1}{N}\omega^2-\frac{1}{N(u'V^{-1}(H)u)}\longrightarrow\omega^2 \quad\text{as}\quad N\longrightarrow\infty.
\end{aligned}$$

Applying the strong law of large numbers and the continuous mapping theorem for almost sure convergence, we get

$$\begin{aligned}
\widehat{\omega^2} &= \frac{1}{N}\sum_{i=1}^{N}\xi_i(H)^2-\left(\frac{1}{N}\sum_{i=1}^{N}\xi_i(H)\right)^2-\frac{1}{u'V^{-1}(H)u}\\
&\overset{\mathbb{P}-as}{\Longrightarrow} \mathbb{E}(\xi_1(H))^2-\mathbb{E}^2(\xi_1(H))-\frac{1}{u'V^{-1}(H)u}=\mathbb{V}ar(\xi_1(H))-\frac{1}{u'V^{-1}(H)u}\\
&= \mathbb{V}ar\left(\phi_1+\frac{u'V^{-1}(H)\epsilon^1}{u'V^{-1}(H)u}\right)-\frac{1}{u'V^{-1}(H)u}\\
&= \mathbb{V}ar\phi_1+\mathbb{V}ar\left(\frac{u'V^{-1}(H)\epsilon^1}{u'V^{-1}(H)u}\right)-\frac{1}{u'V^{-1}(H)u}\\
&= \omega^2+\mathbb{E}\left(\frac{u'V^{-1}(H)\epsilon^1}{u'V^{-1}(H)u}\right)^2-\frac{1}{u'V^{-1}(H)u}=\omega^2.
\end{aligned}$$

Similar computations lead to

$$\begin{aligned}\mathbb{V}ar(\widehat{\omega^2}) &= \frac{N-1}{N^3}\left((N-1)\mathbb{E}(\xi_1(H)-\mu)^4-(N-3)\beta^2\right)\\ &= \frac{2(N-1)}{N^2}\beta^2,\end{aligned}$$

where $\beta = \mathbb{V}ar(\xi_1(H)) = \omega^2 + \dfrac{1}{u'V^{-1}(H)u}$. In the last equality we used the fact that $(\xi_1(H)-\mu)$ is a centered Gaussian with variance β. □

For the case of continuous observation with horizon T, we propose the following estimator $\widetilde{\mu}(T,N)$ defined by

$$\widetilde{\mu}(T,N) = \frac{1}{NT}\sum_{i=1}^{N} Y^i(T).$$

It is easy to see that $\mathbb{E}\left|\dfrac{1}{T}Y^i(T)-\phi_i\right|^2 \leq \dfrac{1}{T^{2-2H}} \longrightarrow 0$, as $T \longrightarrow \infty$ and $\widetilde{\mu}(T,N)$ is consistent as $T,N\to\infty$. The reason we choose this double asymptotic framework, is that we proceed in two steps; in the first step we estimate random effects ϕ_i as the horizon T increases to ∞, then we use the empirical mean and variance to estimate $\theta=(\mu,\omega^2)$, where the random effects are replaced by their estimators.

Theorem 5.2.5. *The estimators $\widehat{\mu}$ and $\widehat{\omega^2}$ are asymptotically normal, i.e.*

$$\sqrt{N}\left(\widehat{\mu}-\mu\right) \overset{\mathcal{L}}{\Longrightarrow} \mathcal{N}\left(0,\omega^2+\frac{1}{u'V^{-1}(H)u}\right),\ \text{as } N\longrightarrow\infty, \tag{5.10}$$

and

$$\sqrt{\frac{N}{2}}\left(\widehat{\omega^2}-\omega^2\right) \overset{\mathcal{L}}{\Longrightarrow} \mathcal{N}\left(0,\left(\omega^2+\frac{1}{u'V^{-1}(H)u}\right)^2\right),\ \text{as } N\longrightarrow\infty. \tag{5.11}$$

Proof. Since $\widehat{\mu}$ is the average of N i.i.d random variables with finite mean and finite variance, (5.10) follows imediately from the central limit theorem (see, [101, Theorem B.97]). Let $\widetilde{\xi}_i(H) = \sqrt{\dfrac{N}{N-1}}(\xi_i(H)-\widehat{\mu})$, $i=1,\cdots,N$ and set $\beta=\omega^2+\dfrac{1}{u'V^{-1}(H)u}$. $\left(\widetilde{\xi}_i(H)\ ,\ i=1,2,\cdots\right)$ is centered Gaussian process, with $\mathbb{E}\widetilde{\xi}_i(H)=0$, $\mathbb{V}ar(\widetilde{\xi}_i(H)) = \mathbb{E}(\widetilde{\xi}_i(H)^2)=\beta$, and $\mathbb{V}ar(\widetilde{\xi}_i(H)^2)=2\beta^2$. By strong law of large numbers, we have $\widetilde{\omega^2} = \dfrac{1}{N}\displaystyle\sum_{i=1}^{N}\widetilde{\xi}_i(H)^2 \overset{\mathbb{P}-as}{\Longrightarrow} \beta$, and the central limit theorem yields $\sqrt{N}\left(\widetilde{\omega^2}-\beta\right) \overset{\mathcal{L}}{\Longrightarrow} \mathcal{N}(0,2\beta^2)$. Since $\sqrt{\dfrac{N}{2}}\left(\widehat{\omega^2}-\omega^2\right) = \alpha_N\sqrt{N}\left(\widetilde{\omega^2}-\beta\right)-\varepsilon_N$, where $\alpha_N = \dfrac{N-1}{\sqrt{2N}}$ and $\varepsilon_N = \dfrac{\beta}{\sqrt{2N}}$, then, using Slutsky theorem, the convergence (5.11) is easily concluded. □

5.2.5 Simulations

We will implement the two population parameter estimators for the model that we have studied to show their empirical behaviour. We will simulate the observed vectors Y^i using (5.3) for two numbers of subjects $N = 50$ and $N = 500$ with different lengths of observations per subject; $n = 2^2$, $n = 2^5$ and $n = 2^8$. The fractional Brownian motions are simulated as in [67]. The experiment is as follows : we set H equal to 0.15, 0.5 and 0.85. For each case, replications involving 400 samples are obtained by resampling n trajectories of Y^i.

The averages of the estimators and their exact against empirical standard deviations are reported in the Tables 5.1-5.3. The tables show that the parameter estimations are generally much closer to their true values as the number of subjects increases. Figures 5.1-5.3 display the histograms densities of the estimators, which reveal the convergence toward a limit distribution also as N is sufficiently large, this confirms what was established before. Looking at Table 5.1, we see that the estimating for ω^2 is not really close to exact values when there are very few observations ($n \leq 2^3$) per subject when $H = 0.85$, this case has been observed every time when H becomes large than $1/2$. In this situation, for the real cases where the true value of ω^2 is not available, it will be better to choose n as large as possible ($n \geq 2^4$) but this leads to huge computational cost for large values of N. Yet, to keep the balance between the computational cost and goodness of fit, a small values of n and sufficiently large values of N should be considered.

True values	$\boldsymbol{H = 0.15}$	$\boldsymbol{H = 0.50}$	$\boldsymbol{H = 0.85}$
$\boldsymbol{N = 50}$	**Mean (Std. dev.'s)**	**Mean (Std. dev.'s)**	**Mean (Std. dev.'s)**
$\mu = -2$	-1.9902 (0.1456 0.1430)	-1.9964 (0.1549 0.1594)	-1.9820 (0.1795 0.2009)
$\omega^2 = 1$	0.9744 (0.2099 0.1942)	1.0303 (0.2376 0.2494)	1.3314 (0.3191 0.3891)
$\boldsymbol{N = 500}$			
$\mu = -2$	-2.0009 (0.0460 0.0441)	-1.9986 (0.0490 0.0515)	-1.9985 (0.0568 0.0634)
$\omega^2 = 1$	0.9964 (0.0670 0.0689)	1.0442 (0.0758 0.0836)	1.2022 (0.1018 0.1228)

Table 5.1. The means with exact (red) and empirical (blue) standard deviations of estimators $\widehat{\mu}$, $\widehat{\omega^2}$ based on 400 samples, with true values $(\mu_0, \omega_0^2) = (-2, 1)$, $\boldsymbol{(T, n) = (5, 2^2)}$, and different values of N $(= 50; 500)$.

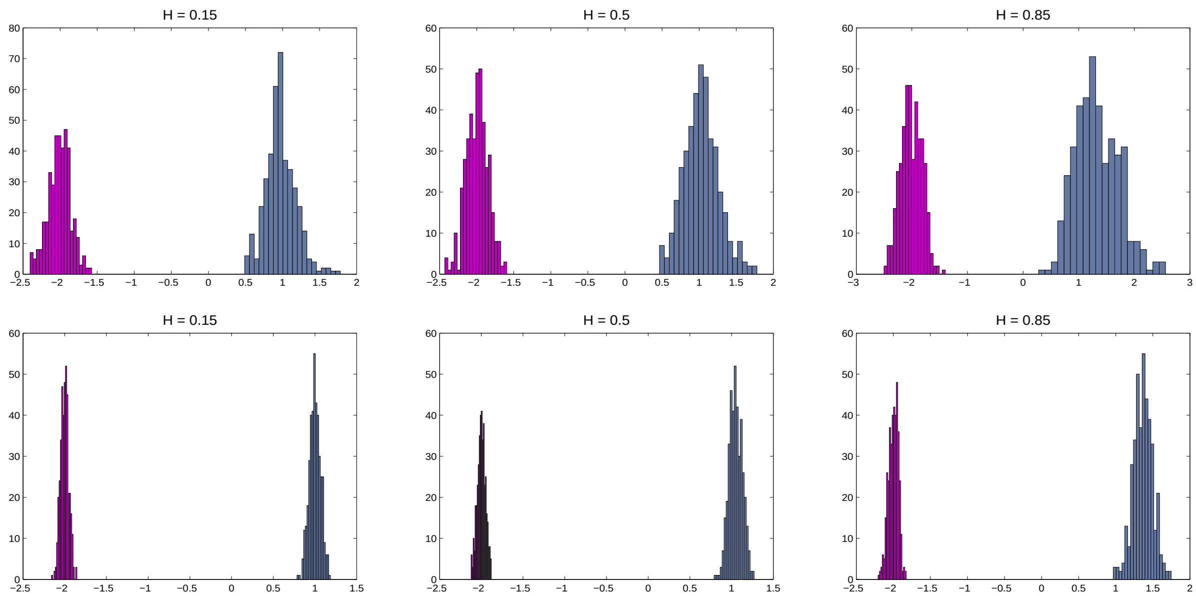

Figure 5.1. Frequency histograms of population parameter estimates based on 400 samples for different values of (N, H). In each box of the two rows (top $N = 50$ and bottom $N = 500$) histograms of $\widehat{\mu}$ (in pink) and $\widehat{\omega^2}$ (in gray) are given for fixed parameters $(\mu, \omega^2, T, \boldsymbol{n}) = (-2, 1, 5, \mathbf{2^2})$.

True values	$\boldsymbol{H = 0.15}$	$\boldsymbol{H = 0.50}$	$\boldsymbol{H = 0.85}$
$\boldsymbol{N = 50}$	**Mean (Std. dev.'s)**	**Mean (Std. dev.'s)**	**Mean (Std. dev.'s)**
$\mu = -2$	-2.0050 (0.1449 0.1427)	-2.0146 (0.1549 0.1518)	-1.9824 (0.1793 0.1920)
$\omega^2 = 1$	0.9713 (0.2077 0.2075)	1.0028 (0.2376 0.2247)	1.0871 (0.3181 0.3391)
$\boldsymbol{N = 500}$			
$\mu = -2$	-2.0057 (0.0458 0.0434)	-1.9979 (0.0490 0.0498)	-2.0038 (0.0567 0.0596)
$\omega^2 = 1$	1.0005 (0.0663 0.0671)	1.0021 (0.0758 0.0758)	1.0849 (0.1015 0.1011)

Table 5.2. The means with exact (red) and empirical (blue) standard deviations of estimators $\widehat{\mu}$, $\widehat{\omega^2}$ based on 400 samples, with true values $(\mu_0, \omega_0^2) = (-2, 1)$, $\boldsymbol{(T, n) = (5, 2^5)}$, and different values of N ($= 50; 500$).

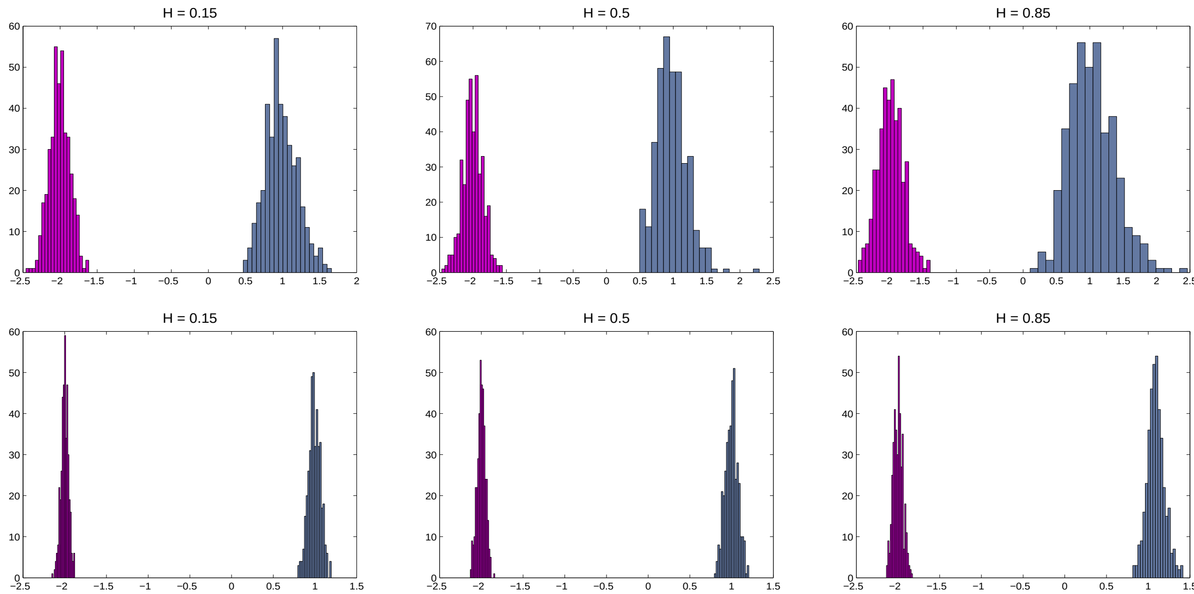

Figure 5.2. Frequency histograms of population parameter estimates based on 400 samples for different values of (N, H). In each box of the two rows (top $N = 50$ and bottom $N = 500$) histograms of $\widehat{\mu}$ (in pink) and $\widehat{\omega^2}$ (in gray) are given for fixed parameters $(\mu, \omega^2, T, \boldsymbol{n}) = (-2, 1, 5, \mathbf{2^5})$.

True values	$\boldsymbol{H = 0.15}$	$\boldsymbol{H = 0.50}$	$\boldsymbol{H = 0.85}$
$\boldsymbol{N = 50}$	**Mean (Std. dev.'s)**	**Mean (Std. dev.'s)**	**Mean (Std. dev.'s)**
$\mu = -2$	-2.0015 (0.1447 0.1454)	-1.9960 (0.1549 0.1563)	-2.0055 (0.1792 0.1709)
$\omega^2 = 1$	0.9996 (0.2073 0.2008)	0.9764 (0.2376 0.2448)	0.9971 (0.3180 0.3323)
$\boldsymbol{N = 500}$			
$\mu = -2$	-1.9997 (0.0458 0.0442)	-2.0009 (0.0490 0.0471)	-2.0006 (0.0567 0.0566)
$\omega^2 = 1$	0.9971 (0.0662 0.0650)	0.9993 (0.0758 0.0747)	1.0083 (0.1015 0.1045)

Table 5.3. The means with exact (red) and empirical (blue) standard deviations of estimators $\widehat{\mu}$, $\widehat{\omega^2}$ based on 400 samples, with true values $(\mu_0, \omega_0^2) = (-2, 1)$, $\boldsymbol{(T, n) = (5, 2^8)}$, and different values of N $(= 50; 500)$.

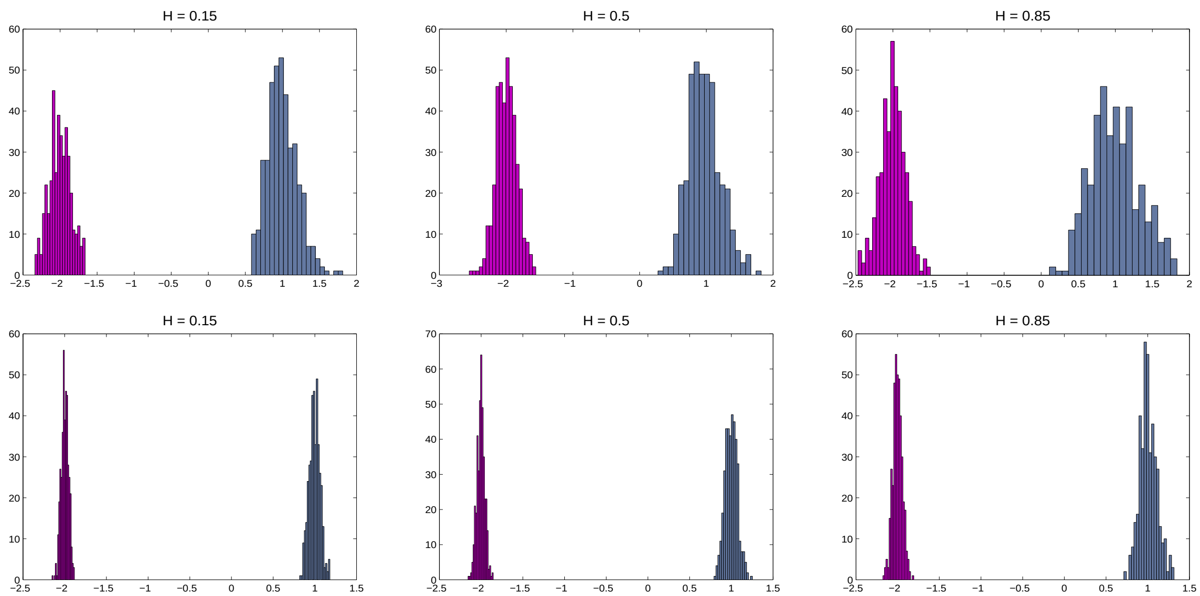

Figure 5.3. Frequency histograms of population parameter estimates based on 400 samples for different values of (N, H). In each box of the two rows (top $N = 50$ and bottom $N = 500$) histograms of $\widehat{\mu}$ (in pink) and $\widehat{\omega^2}$ (in gray) are given for fixed parameters $(\mu, \omega^2, T, \boldsymbol{n}) = (-2, 1, 5, \mathbf{2^8})$.

5.3 NON PARAMETRIC ESTIMATION FOR FSDE'S WITH RANDOM EFFECTS

In this section, we propose a non-parametric estimation for a class of FSDE's with random effects. We precisely consider FSDE's given in (5.1) with $b(t,x) = b(t)$ and $\sigma(t,x) = \sigma(t)$. We build ordinary kernel estimators and histogram estimators and study their L^p-risk ($p = 1$ or 2), when $H > 1/2$. Asymptotic results are evaluated as both $T = T(N)$ and N tend to infinity.

5.3.1 Ordinary Kernel Density Estimators

It is well known that standard kernel density estimators for the unknown density f of ϕ_i are given by

$$\widehat{f}_h(x) = \frac{1}{Nh}\sum_{i=1}^{N} K\left(\frac{x-\phi_i}{h}\right), \quad h > 0, \tag{5.12}$$

where K is an integrable kernel that has to satisfy some regularity conditions on f. The random effects ϕ_i are not observed; it is then natural to replace them by their estimates and prove the consistency of the resulting kernel estimators. We introduce some statistics which have a central role in the estimation procedure. For $i = 1, \cdots, N$, we denote

$$\begin{aligned} U_t^{(1,i)} &= \int_0^t \frac{b(s)}{\sigma^2(s)} dX^i(s), \quad U_t^{(2)} = \int_0^t \frac{b^2(s)}{\sigma^2(s)} ds, \\ R_t^{(i)} &= \int_0^t \frac{a(X^i(s))b(s)}{\sigma^2(s)} ds \quad \text{and} \quad V_t^{(i)} = \int_0^t \frac{b(s)}{\sigma(s)} dW^{H,i}(s). \end{aligned}$$

We know that $V^{(i)} = \left(V_t^{(i)},\ t \geq 0\right)$, $i = 1, \cdots, N$ are Wiener integrals with respect to fBm. A sufficient condition (see, [94, 80]) for the integrals $V^{(i)}$ to be well-defined is that $b(\cdot)/\sigma(\cdot) \in L^2(\mathbb{R}_+) \cap L^1(\mathbb{R}_+)$. The following assumptions are needed to estimate the random effects ϕ_i:

$\mathbf{A}_1$ **:** There exist $c_0, c_1 > 0$ such that

$$c_0^2 \leq \frac{b^2(s)}{\sigma^2(s)} \leq c_1^2, \quad \text{for all } s \in \mathbb{R}_+.$$

$\mathbf{A}_2$ **:** For $i = 1, \cdots, N$, $M_i := \mathbb{E}\left(\int_0^\infty \frac{a^2(X^i(s))}{\sigma^2(s)} ds\right)^2 < \infty.$

Proposition 5.3.1. *Let the assumptions $\boldsymbol{A}_1$-$\boldsymbol{A}_2$ be fulfilled. For $i = 1, \cdots, N$ and $H > 1/2$, we have*

$$\mathbb{E}\left|\widehat{\phi}_{i,T} - \phi_i\right|^2 \longrightarrow 0 \textit{ as } T \to \infty, \textit{ where } \widehat{\phi}_{i,T} := U_T^{(1,i)}/U_T^{(2)}.$$

Proof. Equation (5.1) yields

$$U_t^{(1,i)} = R_t^{(i)} + \phi_i U_t^{(2)} + V_t^{(i)},\ t \le T,\ i = 1, \cdots, N.$$

Thus

$$\frac{1}{2}\mathbb{E}\left|\widehat{\phi}_{i,T} - \phi_i\right|^2 \le \mathbb{E}\left(\frac{R_T^{(i)}}{U_T^{(2)}}\right)^2 + \mathbb{E}\left(\frac{V_T^{(i)}}{U_T^{(2)}}\right)^2. \tag{5.13}$$

We shall show that the expectations on the RHS in (5.13) vanish as T tends to infinity. Applying results in [77, Theorem 1.1] and the Jensen inequality, respectively, we obtain

$$\begin{aligned}
\mathbb{E}\left(\frac{V_T^{(1)}}{U_T^{(2)}}\right)^2 &= \frac{1}{c_0^4 T^2}\mathbb{E}\left(\int_0^T \frac{b(s)}{\sigma(s)} dW^{H,1}(s)\right)^2 \\
&= \frac{C_H^2 c_1^2}{c_0^4 T^2}\left(\int_0^T \left|\frac{b(s)}{\sigma(s)}\right|^{1/H} ds\right)^{2H} \\
&= \frac{C_H^2 c_1^2}{c_0^4 T^{2-2H}} \longrightarrow 0 \text{ as } T \to \infty,
\end{aligned}$$

where C_H is a nonnegative constant due to the Hardy-Littlewood theorem (see, [80]). Using the fact that $|uv| \le \frac{1}{2}\left(\varepsilon u^2 + \frac{v^2}{\varepsilon}\right)$ for all $u, v \in \mathbb{R}$ and $\varepsilon > 0$, we have

$$\begin{aligned}
\mathbb{E}\left(\frac{R_T^{(1)}}{U_T^{(2)}}\right)^2 &= \frac{1}{4}\mathbb{E}\left\{\frac{\varepsilon \int_0^T a^2(X^1(s))/\sigma^2(s)ds + \varepsilon^{-1} U_T^{(2)}}{U_T^{(2)}}\right\}^2 \\
&= \frac{1}{2}\left\{\frac{1}{\varepsilon^2} + \frac{\varepsilon^2}{c_0^4 T^2}\mathbb{E}\left(\int_0^\infty a^2(X^1(s))/\sigma^2(s)ds\right)^2\right\}.
\end{aligned}$$

By choosing $\varepsilon = \sqrt{T}$, we get the desired result and the proof of Proposition 5.3.1 is complete. □

Now, substituting ϕ_i by its estimator $\widehat{\phi}_{i,T}$ in (5.12), we obtain the kernel estimators

$$\widehat{f}_h^{(1)}(x) = \frac{1}{Nh}\sum_{i=1}^{N} K\left(\frac{x - \widehat{\phi}_{i,T}}{h}\right). \tag{5.14}$$

Proposition 5.3.2. *Consider Equation (5.1) where $a(\cdot)$ is unknown and consider the estimator $\widehat{f}_h^{(1)}$ given by (5.14). Assume that $\boldsymbol{A}_1$ and $\boldsymbol{A}_2$ are satisfied. If the kernel K is differentiable with $\|K\|^2 + \|K'\|^2 < \infty$, then*

$$\mathbb{E}\left\|\widehat{f}_h^{(1)} - f\right\|^2 \leq 2\|f_h - f\|^2 + \frac{\|K\|^2}{Nh} + \frac{\|K'\|^2}{T^{1-H}h^3}\left(\frac{1}{T^H} + \frac{M_1}{c_0^4 T^H} + \frac{2C_H^2 c_1^2}{c_0^4 T^{1-H}}\right),$$

where $f_h(x) := K_h * f(x) = \frac{1}{h}\int_{\mathbb{R}} K\left(\frac{x-u}{h}\right) f(u)du.$

Proof. Simple computations show that

$$\begin{aligned}
\mathbb{E}\left\|\widehat{f}_h^{(1)} - f\right\|^2 &= \left\|f - \mathbb{E}(\widehat{f}_h^{(1)})\right\|^2 + \mathbb{E}\left(\left\|\widehat{f}_h^{(1)} - \mathbb{E}(\widehat{f}_h^{(1)})\right\|^2\right) \\
&\leq 2\|f - f_h\|^2 + 2\left\|f_h - \mathbb{E}(\widehat{f}_h^{(1)})\right\|^2 + \mathbb{E}\left(\left\|\widehat{f}_h^{(1)} - \mathbb{E}(\widehat{f}_h^{(1)})\right\|^2\right). \quad (5.15)
\end{aligned}$$

To complete the proof, we evaluate the last two terms in (5.15). Set $\eta_{i,T}(x) = K_h(x - \widehat{\phi}_{i,T}) - \mathbb{E}\left(K_h(x - \widehat{\phi}_{i,T})\right)$, where $K_h(u) = \frac{1}{h}K\left(\frac{u}{h}\right)$. $\eta_{i,T}(x)$, $i = 1, \cdots, N$ are i.i.d random variables with $\mathbb{E}[\eta_{1,T}(x)] = 0$, and with a change of variables $\frac{x - \widehat{\phi}_{1,T}}{h} = y$ in the second inequality below, we get

$$\begin{aligned}
\int_{\mathbb{R}} \mathbb{E}(\eta_{1,T}(x))^2 dx &= \int_{\mathbb{R}} \mathrm{Var}\left(K_h(x - \widehat{\phi}_{1,T})\right) dx \\
&\leq \int_{\mathbb{R}} \mathbb{E}\left(K_h(x - \widehat{\phi}_{1,T})\right)^2 dx \\
&\leq \frac{1}{h^2}\mathbb{E}\int_{\mathbb{R}}\left(K\left(\frac{x - \widehat{\phi}_{1,T}}{h}\right)\right)^2 dx \\
&\leq \frac{1}{h}\int_{\mathbb{R}} K^2(y)dy.
\end{aligned}$$

Thus

$$\begin{aligned}
\mathbb{E}\left(\left\|\widehat{f}_h^{(1)} - \mathbb{E}(\widehat{f}_h^{(1)})\right\|^2\right) &= \mathbb{E}\int_{\mathbb{R}}\left(\widehat{f}_h^{(1)}(x) - \mathbb{E}\widehat{f}_h^{(1)}(x)\right)^2 dx \\
&= \frac{1}{N^2}\mathbb{E}\int_{\mathbb{R}}\left(\sum_{i=1}^N \eta_{i,T}(x)\right)^2 dx \\
&= \frac{1}{N}\int_{\mathbb{R}} \mathbb{E}(\eta_{1,T}(x))^2 dx \leq \frac{\|K\|^2}{Nh}.
\end{aligned}$$

There remains to find an upper bound of the middle term in (5.15). First, note that $f_h(x) = \int_{\mathbb{R}} f(y)K_h(x-y)dy = \mathbb{E}(K_h(x - \phi_1))$. Taylor's theorem with integral remainder yields

$$K_h(x-\widehat{\phi}_{1,T}) - K_h(x-\phi_1) = \frac{(\phi_1-\widehat{\phi}_{1,T})}{h^2}\int_0^1 K'\left(\frac{1}{h}(x-\phi_1+u(\phi_1-\widehat{\phi}_{1,T}))\right)du.$$

Now, set $g(x,u) = K'\left(\frac{1}{h}(x-\phi_1+u(\phi_1-\widehat{\phi}_{1,T}))\right)$, then

$$\begin{aligned}
\left\|f_h - \mathbb{E}(\widehat{f}_h^{(1)})\right\|^2 &= \int_{\mathbb{R}} \left[\mathbb{E}\left(K_h(x-\widehat{\phi}_{1,T}) - K_h(x-\phi_1)\right)\right]^2 dx \\
&\leq \int_{\mathbb{R}} \mathbb{E}\left(K_h(x-\widehat{\phi}_{1,T}) - K_h(x-\phi_1)\right)^2 dx \\
&\leq \mathbb{E}\left[\frac{(\phi_1-\widehat{\phi}_{1,T})^2}{h^4}\int_{\mathbb{R}}\left[\int_0^1 g(x,u)du\right]^2 dx\right] \\
&\leq \mathbb{E}\left[\frac{(\phi_1-\widehat{\phi}_{1,T})^2}{h^4}\left[\int_0^1\left(\int_{\mathbb{R}} g^2(x,u)dx\right)^{1/2} du\right]^2\right].
\end{aligned}$$

The last inequality given above is justified by the generalized Minkowski inequality (see, [110, Lemma A.1]). By change of variables $y = \frac{1}{h}\left(x-\phi_1+u(\phi_1-\widehat{\phi}_{1,T})\right)$, we get $\int_{\mathbb{R}} g^2(x,u)dx = \|K'\|^2 h$. Thus $\left\|f_h - \mathbb{E}(\widehat{f}_h^{(1)})\right\|^2 \leq \frac{\|K'\|^2}{h^3}\mathbb{E}(\phi_1-\widehat{\phi}_{1,T})^2$, which completes the proof (see the proof of Proposition 5.3.1). □

We recall that a kernel of order $l \geq 1$ (for the construction of such a kernel we refer to [110, p.10]) satisfies $\int_{\mathbb{R}} K(u)du = 1$ and $\int_{\mathbb{R}} u^j K(u)du = 0$, for $j = 1, \cdots, l$. For constants $\beta > 0$ and $L > 0$, we define the Nikol'ski class $\mathcal{N}^*(\beta, L)$ as the set of functions $f : \mathbb{R} \longrightarrow \mathbb{R}$, whose derivatives $f^{(l)}$ of order $l = \lfloor\beta\rfloor$ exist and satisfy

$$\left[\int_{\mathbb{R}}\left(f^{(l)}(x+t) - f^{(l)}(x)\right)^2 dx\right]^{1/2} \leq L\,|t|^{\beta-l},\ \forall t \in \mathbb{R},$$

where $\lfloor\beta\rfloor$ denotes the greatest integer strictly less than the real number β.

Corollary 5.3.3. *Assume that $f \in \mathcal{N}^*(\beta, L)$ and that the kernel K has order $l = \lfloor\beta\rfloor$ with $\int_{\mathbb{R}} |u|^\beta |K(u)|\, du < \infty$. Fix $\alpha > 0$ and take $h = \alpha N^{-1/(2\beta+1)}$ and $T^{1-H} \geq N^{(2\beta+3)/(2\beta+1)}$. Then for any $N \geq 1$, the kernel estimator $\widehat{f}_h^{(1)}$ satisfies $\mathbb{E}\left\|\widehat{f}_h^{(1)} - f\right\|^2 \lesssim N^{-2\beta/(2\beta+1)}$.*

Corollary 5.3.4. *Consider Equation (5.1) where $a(\cdot)$ is known. We introduce the estimators*

$$\widehat{f}_h^{(2)}(x) = \frac{1}{Nh}\sum_{i=1}^N K\left(\frac{x-\widetilde{\phi}_{i,T}}{h}\right),$$

where $\widetilde{\phi}_{i,T} := \widehat{\phi}_{i,T} - R_T^{(i)}/U_T^{(2)}$*. Under the assumption* $\boldsymbol{A}_1$*, the estimators* $\widehat{f}_h^{(2)}$ *are consistent with the same optimal rate as for* $\widehat{f}_h^{(1)}$*.*

Remark. The assumption $\mathbf{A}_2$ can be weakened as follows

A'$_2$ **:** For each i, there exists $\delta > 0$ such that

$$\limsup_{t\to\infty} \frac{1}{t^{2-\delta}\log(t)} \mathbb{E}\left(\int_0^t \frac{a^2(X^i(s))}{\sigma^2(s)} ds\right)^2 < \infty.$$

5.3.2 Histogram Estimators

Consider a sequence of partitions of $\mathbb{R}$ of the form $\mathcal{P}_N = \{A_{Nj},\ j = 1, 2, \cdots\}$, $N \geq 1$, where all A_{Nj}'s are Borel sets with finite nonzero Lebesgue measure. We assume that the sequence of partitions is rich enough such that the class of Borel sets $\mathcal{B}$ is equal to

$$\bigcap_{N=1}^{\infty} \boldsymbol{\sigma}\left(\bigcup_{m=N}^{\infty} \mathcal{P}_m\right),$$

where we the symbol $\boldsymbol{\sigma}$ stands for the σ-algebra generated by a class of sets.

Given a sequence of i.i.d random variables $X_1, \cdots, X_N$, with common density f, the histogram estimate is (as in [27]) defined by

$$\mathcal{T}(X.)(x) = \frac{1}{N}\sum_{i=1}^{N} \frac{\chi_{(X_i \in A_{Nj})}}{\lambda(A_{Nj})}, \quad x \in A_{Nj},$$

where λ denotes the Lebesgue measure. For our case, we will consider the following histogram estimators $\widehat{f}_h^{(3)}(x) = \mathcal{T}(\widehat{\phi}.)(x)$; $\widehat{f}_h^{(4)}(x) = \mathcal{T}(\widetilde{\phi}.)(x)$. If the density f of the random effects ϕ_i has compact support, then a good estimator should have compact support as well. To guarantee such property we trim the proposed estimators by $\chi_{\text{supp}f}$.

Let $\mathcal{F}'_b$ denote the class of functions satisfying

(i) f is absolutely continuous with derivative f' (almost everywhere);

(ii) f' is bounded and continuous ($\int_{\mathbb{R}} |f'| < \infty$).

We consider the partitions $A_{Nj} = [hj, h(j+1))$, $j \in \mathbb{Z}$. The following special functions will be used later: $r_N(x) = \frac{x}{h} - j$, $z_N(x) = (1 - 2r_N(x))\, f'(x)$ and

$$\Psi(u) = \sqrt{\frac{2}{\pi}}\left(u\int_0^u e^{-x^2/2}dx + e^{-u^2/2}\right), \quad u \geq 0.$$

Proposition 5.3.5. *Let $f \in \mathcal{F}'_b$ have compact support A and assume that $1, \cdots, J$ are nonzero indices for which $\lambda\left(A_{Nj} \cap A\right) \neq 0$ and $T = T(N)$, where λ is the Lebesgue measure. Then, the following statements hold true:*

(i) *When $a(\cdot)$ is unknown, under the assumptions $\boldsymbol{A}_1$ and $\boldsymbol{A}_2$, we have*

$$\mathbb{E}\left\|\widehat{f}_h^{(3)} - f\right\|_1 \leq \psi_1(N,h) + \psi_2(h) + \frac{dJ}{h^2\sqrt{T^{1-H}}} + o\left(h + \frac{1}{\sqrt{Nh}}\right),$$

where d is some nonnegative constant and

$$\psi_1(N,h) = \int_{\mathbb{R}} \sqrt{\frac{f}{Nh}}\Psi\left(\frac{h}{2}\left|z_N\right|\sqrt{\frac{Nh}{f}}\right) \to 0 \ \text{as } h \to 0, \ Nh \to \infty,$$

$$\psi_2(h) = \frac{2}{N}\sum_{i=1}^{N}\sum_{j=1}^{J}\mathbb{P}\left(\phi_i \in A_{Nj}\right)^{1/2} \to 0 \ \text{as } h \to 0 \ \text{(see Lemma 6.0.12)}.$$

(ii) *When $a(\cdot)$ is known, we may relax the assumption $\boldsymbol{A}_2$, and the same result holds for $\widehat{f}_h^{(4)}$.*

Proof. By virtue of [27, Theorem 6], and for sufficiently small h such that $Nh \to \infty$, we have

$$\begin{aligned}
\mathbb{E}\left\|\widehat{f}_h^{(3)} - f\right\|_1 &\leq \sum_j \mathbb{E}\int_{A_{Nj}\cap A}\left|\widehat{f}_h^{(3)}(x) - f(x)\right| dx \\
&\leq \sum_j \mathbb{E}\int_{A_{Nj}}\left|\mathcal{T}(\phi.)(x) - f(x)\right| dx + \sum_{j\leq J}\mathbb{E}\int_{A_{Nj}}\left|\mathcal{T}(\widehat{\phi}.)(x) - \mathcal{T}(\phi.)(x)\right| dx \\
&\leq \mathbb{E}\left\|\mathcal{T}(\widehat{\phi}.) - f\right\|_1 + \frac{1}{Nh}\sum_{j=1}^{J}\sum_{i=1}^{N}\int_{hj}^{h(j+1)}\mathbb{E}\left|\chi_{(\widehat{\phi}_{i,T}\in A_{Nj})} - \chi_{(\phi_i\in A_{Nj})}\right| dx \\
&\leq \psi_1(N,h) + o\left(h + \frac{1}{\sqrt{Nh}}\right) \\
&\qquad + \frac{1}{Nh}\sum_{j=1}^{J}\sum_{i=1}^{N}\int_{hj}^{h(j+1)}\mathbb{E}\left|\chi_{(\widehat{\phi}_{i,T}\in A_{Nj})} - \chi_{(\phi_i\in A_{Nj})}\right| dx.
\end{aligned}$$

Let $\nu(N,J,h)$ denote the last term in the last inequality above. The sequence $\widehat{\phi}_{i,T(N)}$ converges weakly to ϕ_i, since it converges in L^2-sense as N tends to infinity (say $T(N) \to \infty$). Thus, by using Lemma 6.0.11, we obtain

$$\nu(N,J,h) \leq \frac{\sqrt{2}}{N}\sum_{j=1}^{J}\sum_{i=1}^{N}\mathbb{P}(\phi_i \in A_{Nj})^{1/2}\left[\mathbb{P}(\widehat{\phi}_{i,T} \notin A_{Nj})^{1/2} + \mathbb{P}(\phi_i \notin A_{Nj})^{1/2}\right].$$

Let $\alpha \in (0,1)$ to be specified later. We apply Lemma 6.0.10 to get

$$\begin{aligned}\mathbb{P}\left(\widehat{\phi}_{i,T} \notin A_{Nj}\right) &\leq \mathbb{P}\left(\left|\widehat{\phi}_{i,T} - h(j+1/2)\right| \geq h/2\right) \\ &\leq \mathbb{P}\left(\left|\widehat{\phi}_{i,T} - \phi_i\right| \geq (1-\alpha)h/2\right) + \mathbb{P}\left(|\phi_i - h(j+1/2)| \geq \alpha h/2\right) \\ &\leq \frac{4\mathbb{E}\left(\widehat{\phi}_{i,T} - \phi_i\right)^2}{(1-\alpha)^2 h^2} + \mathbb{P}\left(\phi_i \notin A_{Nj}^{(\alpha)}\right) \leq \frac{d_1}{(1-\alpha)^2 h^2 T^{1-H}} + 1,\end{aligned}$$

where d_1 is some nonnegative constant (see the proof of Proposition 5.3.1) and $A_{Nj}^{(\alpha)} = \left(h(j + \frac{1-\alpha}{2}), h(j + \frac{1+\alpha}{2})\right)$. Similarly, one can prove that $\mathbb{P}(\phi_i \notin A_{Nj}) \leq \frac{d_1}{(1-\alpha)^2 h^2 T^{1-H}} + 1$. Thus

$$\begin{aligned}\nu(N,J,h) &\leq \frac{2}{N}\sum_{j=1}^{J}\sum_{i=1}^{N}\left[\mathbb{P}(\phi_i \in A_{Nj})\left(\frac{d_1}{(1-\alpha)^2 h^2 T^{1-H}} + 1\right)\right]^{1/2} \\ &\leq \frac{2\sqrt{d_1}J}{(1-\alpha)h\sqrt{T^{1-H}}} + \frac{2}{N}\sum_{j=1}^{J}\sum_{i=1}^{N}\left[\mathbb{P}(\phi_i \in A_{Nj})\right]^{1/2},\end{aligned}$$

where we used the fact that $\sqrt{u+v} \leq \sqrt{u} + \sqrt{v}$, for all $u, v \in \mathbb{R}_+$. Set $d = 2\sqrt{d_1}$ and $\alpha = 1 - h$ to complete the proof. □

Proposition 5.3.6. *We have*

$$\psi_2(h) = O(h^\delta), \quad \text{where } \delta \in (0, 1/2). \tag{5.16}$$

Proof. Let $\delta, \delta^* \in (0,1)$ such that $\delta + \delta^* = 1$. It is easy to see that

$$\begin{aligned}\mathbb{P}(\phi_i \in A_{Nj}) &= \mathbb{P}(\phi_i \in A_{Nj})^{\delta^*}\left(\int_{hj}^{h(j+1)} f(t)dt\right)^{\delta} \\ &\leq \left[\sup_{i,j}\mathbb{P}(\phi_i \in A_{Nj})\right]^{\delta^*}\sup_t f(t)^\delta h^\delta \\ &\leq \frac{e^{-j\delta^*}}{j!}\sup_t f(t)^\delta h^\delta, \quad h \in (0, h_0),\end{aligned}$$

where h_0 is some nonnegative number independent of i and j. Thereby,

$$\psi_2(h) = \sup_t f(t)^{\delta/2} h^{\delta/2} \sum_{j \geq 1} \frac{e^{-j\delta^*/2}}{\sqrt{j!}} < \infty.$$

□

Let $T = T(N) \geq J^{4/(1-H)}$ so that

$$\frac{dJ}{h^2 \sqrt[4]{T^{1-H}}} = O\left(h^{\delta'}\right), \text{ and set } \ h \propto N^{-\delta''}. \tag{5.17}$$

As mentioned in [27, Theorem 6],

$$\psi_1(N,h) + o\left(h + \frac{1}{\sqrt{Nh}}\right) = O\left(N^{-1/3}\right). \tag{5.18}$$

Fitting rates of convergence given in (5.16),(5.17) and (5.18), we choose $\delta' = \delta$, $\delta'' = 1/(3\delta)$. An arbitrary choice of δ may violate the crucial condition $Nh \longrightarrow \infty$ as $h \to 0$. Choosing $\delta \in (1/3, 1/2)$, we guarantee that all conditions on T, J, N and h are fulfilled. Finally $\mathbb{E}\left\|\widehat{f}_h^{(3)} - f\right\|_1 = O\left(N^{-1/3}\right)$. In a similar fashion, we can prove that $\widehat{f}_h^{(4)}$ as well as $\widehat{f}_h^{(3)}$ have the same rates of convergence .

5.3.3 Numerical Simulations

As an example, we consider the following Langevin equation as dynamics of the subject X^i:

$$\begin{aligned} dX^i(t) &= \left(-\lambda X^i(t) + \phi_i b(t)\right) dt + \sigma dW^{H,i}(t), \ t \leq T \\ X^i(0) &= x^i \in \mathbb{R}, \end{aligned} \tag{5.19}$$

where $H > 1/2$, λ, $\sigma > 0$ and ϕ_i is a random variable such that $\mathbb{E}\left|\phi_i\right|^4 < \infty$, $i = 1, \ldots, N$. Assume that $b_1^2 \leq b(t)^2 \leq b_2^2$, for all $t \leq T$. The common density f of ϕ_i can be estimated by $\widehat{f}_h^{(2)}$ and $\widehat{f}_h^{(4)}$, since the condition $\mathbf{A}_1$ is trivial.

For illustration, we simulate model (5.19) with $b(t) = \sigma = 1$, estimate the densities of the random effects and compare these to the true data-generating density. In detail, we use up to 25 exact simulations with $\lambda = 3 \times 10^{-3}$, $x^i = 0$, $N = 1000$ and $T = 100; 10$. The random effects are Gaussian distributed, $\mathcal{N}(1, 0.8)$, and Gamma distributed, $\Gamma(2, 0.9)$, where 2 is the shape parameter and 0.9 the scale parameter. Figures 5.4, 5.5, 5.6 and 5.7 display the estimates $\widehat{f}_h^{(2)}$ and $\widehat{f}_h^{(4)}$ for different values of the Hurst index, $H \in \{0.25, 0.75, 0.85\}$ and $T = 100; 10$. Improving the accuracy of our estimators requires that both N and T be sufficiently large. However, for T being only moderately large (say $T = 10$) and/or $H < 1/2$ (which is not supported by our theoretical framework), the estimated curves match the theoretical curves satisfyingly well. In general, the estimators $\widehat{f}_h^{(1)}$ and $\widehat{f}_h^{(3)}$ are recommended in the case where $a(\cdot)$ is unknown, but one has to verify the condition $\mathbf{A'}_2$. The estimators $\widehat{f}_h^{(2)}$ and $\widehat{f}_h^{(4)}$ require less assumptions, but the results are more time-consuming as we need to compute $\widehat{\phi}_{i,T}$ and $R_T^{(i)}/U_T^{(2)}$; while $\widehat{f}_h^{(1)}$ and $\widehat{f}_h^{(3)}$ require only $\widehat{\phi}_{i,T}$.

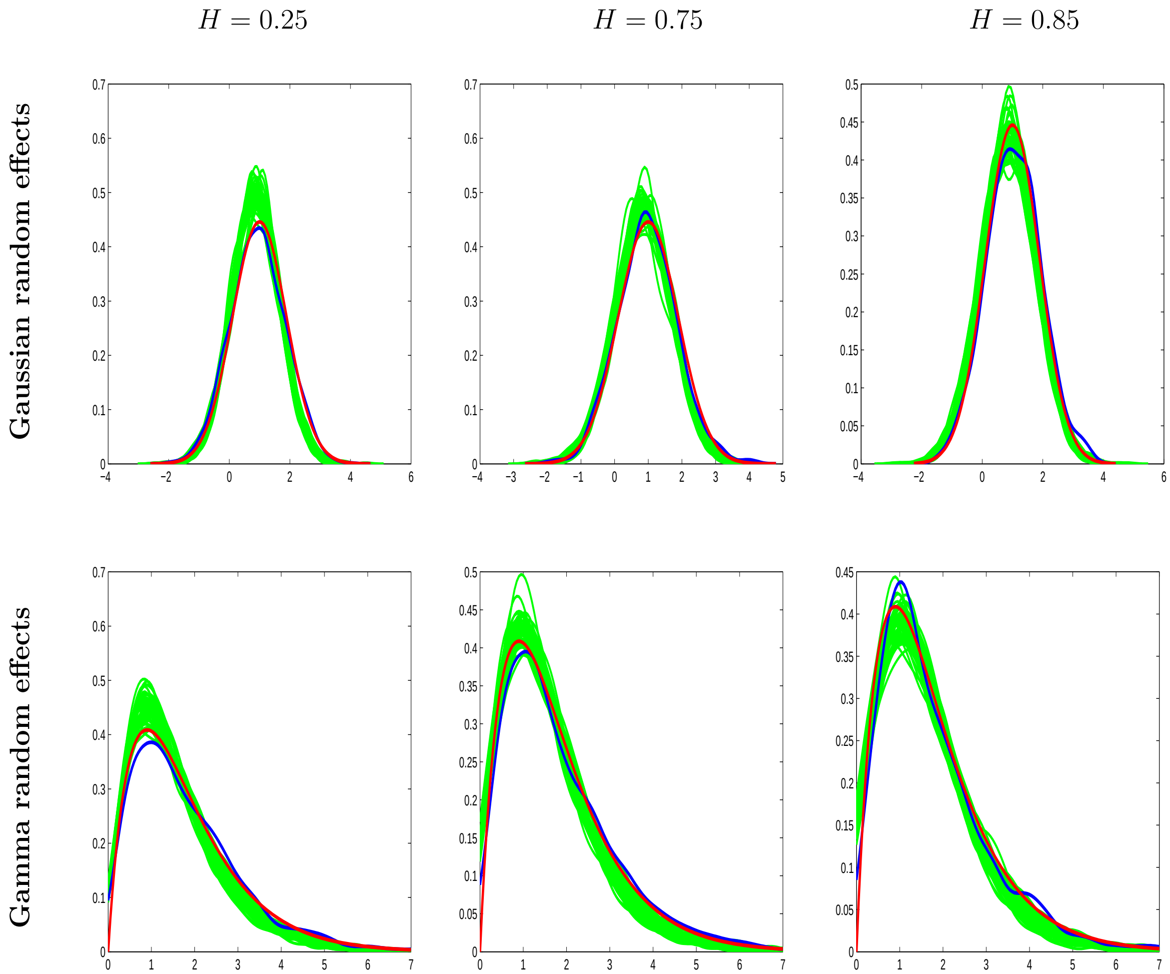

Figure 5.4. **Kernel estimates $\hat{f}_h^{(2)}$ for Ornstein-Uhlenbeck process with additive random effects:** We drew 50 i.i.d. realizations of model (5.19) for each of the following settings. *First row:* Gaussian distributed random effects, *second row:* gamma distributed random effects, *columns:* different values for the Hurst index H. The thin green lines show the 25 kernel estimates $\hat{f}_h^{(2)}$. The true density is shown in bold red, and a standard kernel density estimator for one sample of ϕ_i's (which is unobserved in a real-case scenario) in blue bold. We chose $N = 1000$ and $T = 100$. For more details, see Section 5.3.3.

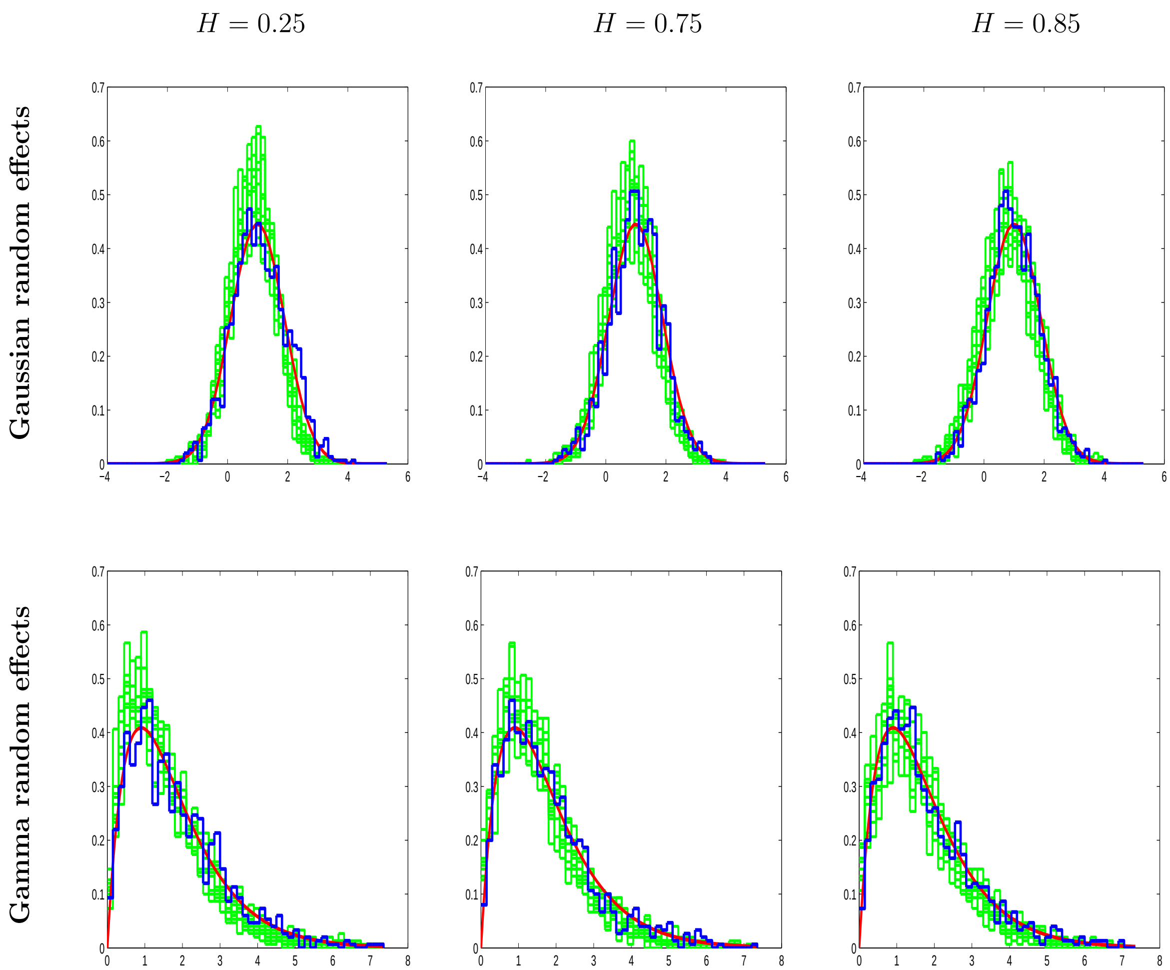

Figure 5.5. **Histogram estimates $\hat{f}_h^{(4)}$ for Ornstein-Uhlenbeck process with additive random effects:** We drew 10 i.i.d. realizations of model (5.19) for each of the following settings. *First row:* Gaussian distributed random effects, *second row:* gamma distributed random effects, *columns:* different values for the Hurst index H. The thin green lines show the 10 histogram estimates $\hat{f}_h^{(4)}$. The true density is shown in bold red, and an exact histogram for one sample of ϕ_i's (which is unobserved in a real-case scenario) in blue bold. We chose $N = 1000$ and $T = 100$. For more details, see Section 5.3.3.

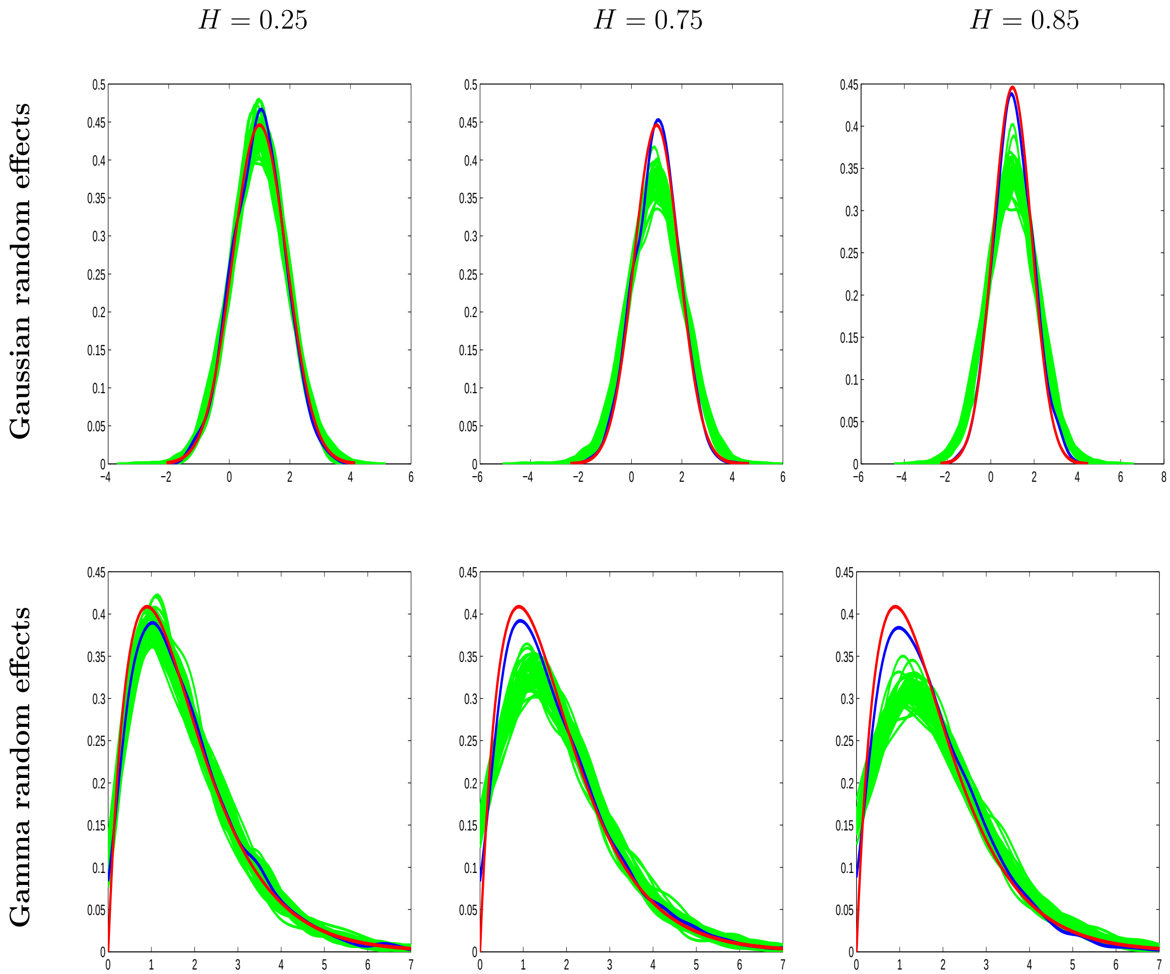

Figure 5.6. **Kernel estimates $\hat{f}_h^{(2)}$ for Ornstein-Uhlenbeck process with additive random effects:** We drew 50 i.i.d. realizations of model (5.19) for each of the following settings. *First row:* Gaussian distributed random effects, *second row:* gamma distributed random effects, *columns:* different values for the Hurst index H. The thin green lines show the 25 kernel estimates $\hat{f}_h^{(2)}$. The true density is shown in bold red, and a standard kernel density estimator for one sample of ϕ_i's (which is unobserved in a real-case scenario) in blue bold. We chose $N = 1000$ and $T = 10$. For more details, see Section 5.3.3.

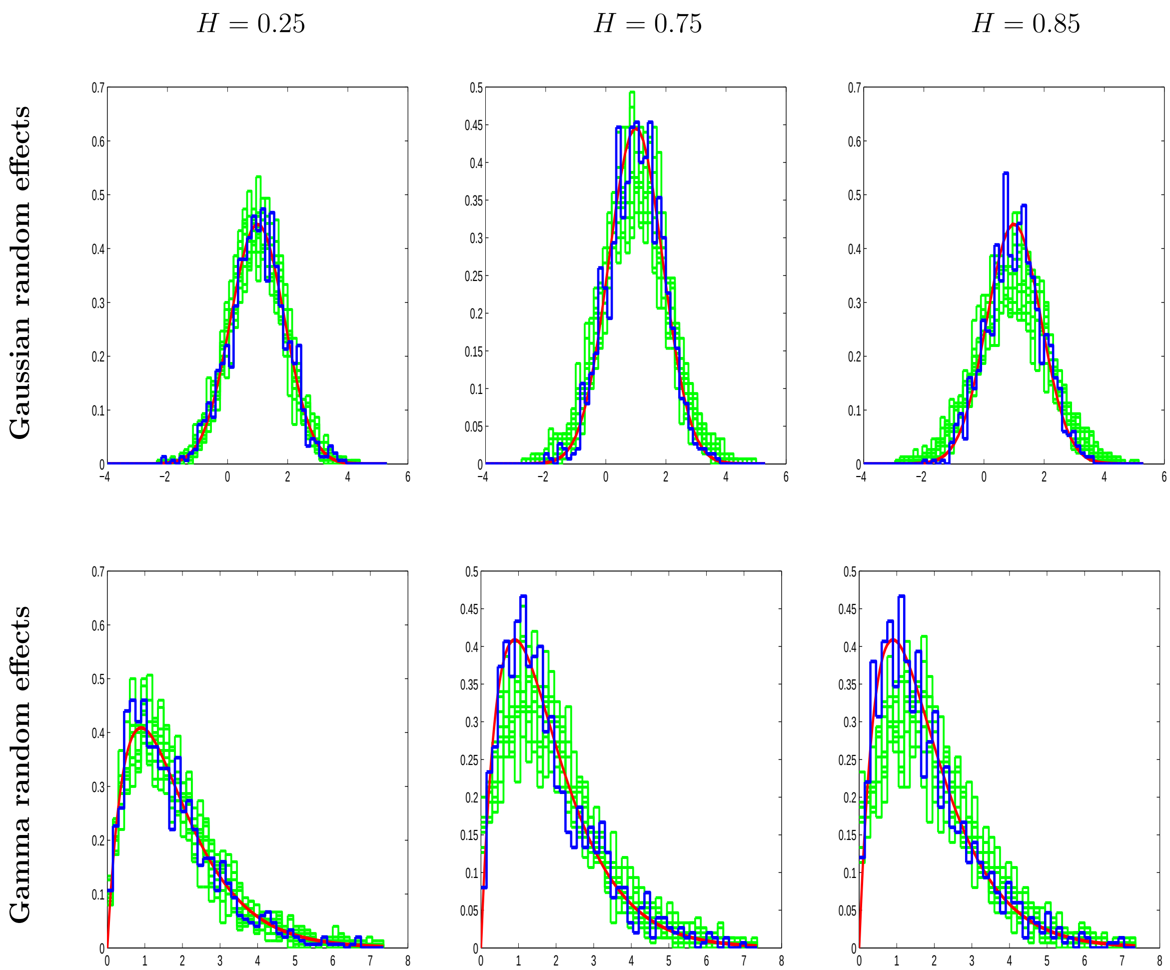

Figure 5.7. **Histogram estimates $\hat{f}_h^{(4)}$ for Ornstein-Uhlenbeck process with additive random effects:** We drew 10 i.i.d. realizations of model (5.19) for each of the following settings. *First row:* Gaussian distributed random effects, *second row:* gamma distributed random effects, *columns:* different values for the Hurst index H. The thin green lines show the 10 histogram estimates $\hat{f}_h^{(4)}$. The true density is shown in bold red, and an exact histogram for one sample of ϕ_i's (which is unobserved in a real-case scenario) in blue bold. We chose $N = 1000$ and $T = 10$. For more details, see Section 5.3.3.

5.4 NON PARAMETRIC ESTIMATION FOR FSDE'S WITH RANDOM EFFECTS AND SMALL FRACTIONAL DIFFUSION

This section deals with the non-parametric estimation problem for processes of type (5.1) with $a(x) = 0$, $b(t,x) = b(x)$ and $\sigma(t,x) = \varepsilon$. We propose a class of estimators of random effects (common) density f, when $H > 1/2$. The asymptotic behaviour of the proposed estimators is established as $\varepsilon \to 0$ and N (the number of subjects) tends to infinity.

5.4.1 Model, Notations and Procedures of Estimation

We are concerned with N subjects $\left(X^i(t), \mathcal{F}_t^i, t \leq T\right)$ with dynamics ruled by the following general linear stochastic differential equations:

$$\begin{aligned} dX^i(t) &= \phi_i b(X^i(t))dt + \varepsilon dW^{i,H}(t),\ 0 \leq t \leq T, \\ X^i(0) &= x_0^i \in \mathbb{R}, \quad i = 1, \cdots, N, \quad H \in (1/2, 1), \end{aligned} \tag{5.20}$$

where $b(\cdot)$ is a known function and the random effects ϕ_i are $\mathcal{F}_0^i$-measurable with common density f to be estimated under some regularity conditions (to be specified later). Assume also that

$$\mu_f(\lambda) := \int e^{\lambda|u|} f(u)du < \infty, \quad \forall \lambda > 0, \tag{5.21}$$

which is obviously satisfied by Gaussian and Beta distributions. The following conditions guarantee the existence and uniqueness of solutions to (5.20)(see, Theorem 2.2.3 or [80, p. 197]).

- A_1: There exists $L > 0$ such that $|ub(x) - u'b(x')| \leq L\left(|x - x'| + |u - u'|\right)$, for all $u, u', x, x' \in \mathbb{R}$
- A_2: There exists a nonnegative constant R such that

$$|ub(x)| \leq R\left(1 + |u| + |x|\right),\ \forall u, x \in \mathbb{R}$$

Consider the differential equations in the limiting system of (5.20), that is, for $\varepsilon = 0$, given by

$$dx^i(t) = \phi_i b(x^i(t))dt, \quad x^i(0) = x_0^i,\ 0 \leq t \leq T. \tag{5.22}$$

Our procedures of estimation will be split into three steps

1. Estimating the functions $\phi_i b(x^i(t))$ for any $t \in [0, T]$ from the observations $X^1, \cdots, X^N$.

2. Estimating the random effects ϕ_i.

3. Estimating the common density function f.

5.4.2 Preliminary Results

We state our main result on the density estimators of the random variable ϕ_i. Namely, the non standard kernel density estimators for the unknown density f of unobserved ϕ_i, and on his asymptotic property. in the sequel, $g_{i,\varepsilon'} = (g_{i,\varepsilon'}(t),\ t \geq 0)$, $i = 1, \cdots, N$ be a processes defined by

$$g_i(t) := g_{i,\varepsilon'}(t) = \begin{cases} X^i(t), \text{ if } \varepsilon' = 1 \\ \\ x^i(t), \text{ if } \varepsilon' = 0. \end{cases}$$

In many proofs below, we focuss on one single subject and simplify notations by omitting indices.

Lemma 5.4.1. *Under the assumptions $\boldsymbol{A}_1$ and $\boldsymbol{A}_2$, the following statements hold*

(i) *There exists a nonnegative constant V such that*

$$\sup_{0 \leq t \leq T} \mathbb{E}(g_i(t)^2) \leq V, \quad \textit{for all } i. \tag{5.23}$$

(ii) *There exists a constant $C > 0$ such that*

$$\mathbb{E}\left|x^i(t) - x^i(s)\right|^2 \leq C\left|t - s\right|^2, \quad \textit{for all } i. \tag{5.24}$$

Proof. Set $g_t = g_1(t)$ and $x_t = x^1(t)$. For the statement (i), we fix $t \in [0, T]$ and set $U_t = \mathbb{E}g_t^2$. Using the assumption $\mathbf{A}_2$ in the third inequality, we obtain

$$\begin{aligned} U_t &= \mathbb{E}\left(x_0 + \phi\int_0^t b(g_s)ds + \varepsilon\varepsilon' W_t^H\right)^2 \\ &\leq 3\left(x_0^2 + (\varepsilon\varepsilon')^2\mathbb{E}(W_t^H)^2\right) + 3\mathbb{E}\left(\phi\int_0^t b(g_s)ds\right)^2 \\ &\leq 3\left(x_0^2 + (\varepsilon\varepsilon')^2 t^{2H}\right) + 3t\int_0^t \mathbb{E}(\phi b(g_s))^2 ds \\ &\leq 3\left(x_0^2 + (\varepsilon\varepsilon')^2 t^{2H}\right) + 9R^2 t\int_0^t (1 + \mathbb{E}(\phi^2) + \mathbb{E}g_s^2)ds \\ &\leq V_t + 9R^2 T\int_0^t U_s ds, \end{aligned}$$

where $V_t = 3\left(x_0^2 + (\varepsilon\varepsilon')^2 t^{2H} + 3R^2(1+\mathbb{E}\phi^2)t^2\right)$. Applying the Gronwall lemma (see, [73, Lemma 4.15]) to get

$$U_t \le V_t + 9R^2T\int_0^t e^{9R^2T(t-s)}V_s ds \le V, \quad \forall t \le T,$$

where $V = V_T + 9R^2T\int_0^T e^{9R^2T(T-s)}V_s ds$. For the statement (ii), let $t > s$. We have

$$\begin{aligned}
\mathbb{E}(x_t - x_s)^2 &= \mathbb{E}\left(\int_s^t \phi b(x_u)du\right)^2 \\
&\le (t-s)\int_s^t \mathbb{E}(\phi b(x_u))^2 du \\
&\le 3R^2(t-s)\int_s^t \left(1+\mathbb{E}\phi^2 + \mathbb{E}x_u^2\right)du \\
&\overset{(5.23)}{\le} 3R^2\left(1+\mathbb{E}\phi^2+V\right)(t-s)^2.
\end{aligned}$$

□

Let $K(u)$ be a bounded function with finite support $[A,B]$ $(A<B)$ and satisfies

$$K(u) = 0, \quad \text{for all } u \notin [A,B] \quad \text{and} \quad \int_A^B K(u)du = 1.$$

It is clear that $\int_{\mathbb{R}} |K(u)|^r du < \infty$, for all $r > 0$. For $i = 1, \cdots, N$ set

$$\widehat{Q}_{i,\varepsilon}(t) := \frac{1}{h}\int_0^T K\left(\frac{s-t}{h}\right) dX^i(s), \quad t \in [0,T].$$

Theorem 5.4.2. *For $i = 1, \cdots, N$*

$$\sup_{0\le t\le T} \mathbb{E}\left|\widehat{Q}_{i,\varepsilon}(t) - \phi_i b(x^i(t))\right|^2 \longrightarrow 0, \tag{5.25}$$

provided that $\varepsilon^2 h^{2H-2} \longrightarrow 0$, as $\varepsilon, h \longrightarrow 0$.
If we assume that $\varepsilon = \varepsilon_n$ and $h = \varepsilon_n^{1/(3-2H)}$, there exists $h_0 > 0$ so that

$$\left\|\widehat{Q}_{i,\varepsilon_n}(t) - \phi_i b(x^i(t))\right\|_{L^2(\Omega)} = O\left(\varepsilon_n^{\frac{1}{3-2H}}\right), \tag{5.26}$$

for all $t \in (0,T)$ and $n \in \mathbb{N}^*$, with $\varepsilon_n < \left(1 \wedge h_0^{3-2H}\right)$.

Proof. Set $\widehat{Q}_{\varepsilon}(t) = \widehat{Q}_{1,\varepsilon_n}(t)$, $\phi = \phi_1$, $W_t^H = W^{1,H}(t)$, $x_t = x^1(t)$ and $X_t = X^1(t)$. For sufficiently small h (say $h < h_0$), we have $[A, B] \subset (-t/h, (T-t)/h)$. Thus

$$\int_{-t/h}^{(T-t)/h} K(u) b\left(X_{t+hu}\right) du = \int_A^B K(u) b\left(X_{t+hu}\right) du, \quad \text{for all } 0 < h < h_0. \tag{5.27}$$

Hence,

$$\begin{aligned}
\mathbb{E}\left(\widehat{Q}_{\varepsilon}(t) - \phi b(x_t)\right)^2 &= \mathbb{E}\left(\frac{1}{h}\int_0^T K\left(\frac{s-t}{h}\right) dX_s - \phi b(x_t)\right)^2 \\
&= \mathbb{E}\left(\frac{1}{h}\int_0^T K\left(\frac{s-t}{h}\right)\left(\phi b(X_s)ds + \varepsilon dW_s^H\right) - \phi b(x_t)\right)^2 \\
&= \mathbb{E}\Bigg(\int_{-t/h}^{(T-t)/h} K(u)(\phi b(X_{t+hu}))du \\
&\qquad - \int_A^B K(u)(\phi b(x_t))du + \frac{\varepsilon}{h}\int_0^T K\left(\frac{s-t}{h}\right) dW_s^H\Bigg)^2 \\
&= \mathbb{E}\Bigg(\int_A^B K(u)\left(\phi b(X_{t+hu}) - \phi b(x_{t+hu})\right) du \\
&\qquad - \int_A^B K(u)(\phi b(x_{t+hu}) - \phi b(x_t))du + \frac{\varepsilon}{h}\int_0^T K\left(\frac{s-t}{h}\right) dW_s^H\Bigg)^2 \\
&= \mathbb{E}\left(I_1 + I_2 + I_3\right)^2 \leq 3\left(\mathbb{E}I_1^2 + \mathbb{E}I_2^2 + \mathbb{E}I_3^2\right),
\end{aligned} \tag{5.28}$$

where I_1, I_2 and I_3 designate the integrals in the fourth equality among those above. We have

$$\begin{aligned}
\mathbb{E}I_1^2 &= \mathbb{E}\left(\int_A^B K(u)\left(\phi b(X_{t+hu}) - \phi b(x_{t+hu})\right) du\right)^2 \\
&\leq (B-A)\int_A^B K^2(u)\mathbb{E}\left(\phi b(X_{t+hu}) - \phi b(x_{t+hu})\right)^2 du \\
&\leq L^2(B-A)\sup_{0\leq s\leq T}\mathbb{E}(X_s - x_s)^2 \int_{\mathbb{R}} K^2(u)du \\
&\leq C(L,T,H,f)\varepsilon^2 L^2(B-A)\int_{\mathbb{R}} K^2(u)du \longrightarrow 0 \text{ as } \varepsilon \longrightarrow 0,
\end{aligned} \tag{5.29}$$

where $C(L, T, H, f)$ is a nonnegative constant due to Lemma 6.0.13,

$$
\begin{aligned}
\mathbb{E}I_2^2 &= \mathbb{E}\left(\int_A^B K(u)(\phi b(x_{t+hu}) - \phi b(x_t))du\right)^2 \\
&\leq \int_A^B K^2(u)du \int_A^B \mathbb{E}(\phi b(x_{t+hu}) - \phi b(x_t))^2 du \text{ (by Cauchy Schwarz inequality)} \\
&\leq L^2 \int_{\mathbb{R}} K^2(u)du \int_A^B \mathbb{E}(x_{t+hu} - x_t)^2 du \text{ (by } A_1 \text{)} \\
&\overset{(5.24)}{\leq} L^2 \int_{\mathbb{R}} K^2(u)du \int_A^B Ch^2 |u|^2 du \\
&\leq C(L, A, B, K)h^2 \longrightarrow 0 \text{ as } h \longrightarrow 0, \qquad (5.30)
\end{aligned}
$$

where $C(L, A, B, K) = CL^2 \int_{\mathbb{R}} K^2(u)du \int_A^B |u|^2 du$, with C is the nonnegative constant appearing in (5.24).

Applying Theorem 2.2.2 and the Jensen inequality, respectively, we obtain

$$
\begin{aligned}
\mathbb{E}I_3^2 &= \mathbb{E}\left(\frac{\varepsilon}{h}\int_0^T K\left(\frac{s-t}{h}\right) dW_s^H\right)^2 \\
&\leq \frac{C_H^2 \varepsilon^2}{h^2}\left(\int_0^T K^{1/H}\left(\frac{s-t}{h}\right) ds\right)^{2H} \\
&\leq \frac{C_H^2 \varepsilon^2 h^{2H}}{h^2}\left(\int_A^B K^{1/H}(u)du\right)^{2H} \\
&\leq C_H^2 \varepsilon^2 h^{2H-2}(B-A)^{2H-1}\int_{\mathbb{R}} K^2(u)du \longrightarrow 0 \text{ as } \varepsilon, h \longrightarrow 0, \qquad (5.31)
\end{aligned}
$$

provided that $\varepsilon^2 h^{2H-2} \longrightarrow 0$. The second inequality above is justified by change of variables; $u = \dfrac{s-t}{h}$. Note that C_H is a nonnegative constant due to the Hardy-Littlewood theorem (see, [80]). As we can see in (5.29), (5.30) and (5.31), the last three expectations in (5.28) are bounded independently of t, thereby (5.25) follows immediately. For the statement (5.26), note that (5.27) is valid if $h < h_0$, that is $\varepsilon_n < h_0^{3-2H}$. Combining (5.29), (5.30) and (5.31) with the condition $\varepsilon = \varepsilon_n < 1$, we get

$$
\begin{aligned}
\left\|\widehat{Q}_{\varepsilon_n}(t) - \phi b(x_t)\right\|_{L^2(\Omega)} &= O\left(\varepsilon_n \vee h \vee \varepsilon_n h^{H-1}\right) \\
&= O\left(\varepsilon_n^{1/(3-2H)}\right), \quad \text{for all } t \in [0, T],
\end{aligned}
$$

which completes the proof. □

The particular cases ($t = 0$ and $t = T$) influence the choice of support $[A, B]$, since (5.27) is only valid if $A = 0$ for the first case and $B = 0$ for the second one. Now, we are ready to estimate the random effects ϕ_i by

$$\widehat{\phi}_{i,n}^{(1)} := \frac{\widehat{Q}_{i,\varepsilon_n}(t)}{b(x^i(t))}, \quad \text{for some fixed } t \in [0,T]. \tag{5.32}$$

It is natural to replace $x^i(t)$ by $X^i(t)$ in (5.32), since $X^i(t) \to x^i(t)$ in L^2-sense, for all t as $\varepsilon_n \to 0$. Thus we propose the following estimators of ϕ_i

$$\widehat{\phi}_{i,n}^{(2)} := \frac{\widehat{Q}_{i,\varepsilon_n}(t)}{b(X^i(t))}, \quad \text{for some fixed } t \in [0,T], \tag{5.33}$$

which converge toward ϕ_i in probability. The following theorem gives the rate of convergence of $\widehat{\phi}_{i,n}^{(1)}$ toward ϕ_i in probability.

Theorem 5.4.3. *For $i = 1, \cdots, N$ and $t \in [0,T]$, we have*

$$\left|\widehat{\phi}_{i,n}^{(1)} - \phi_i\right| = O_{\mathbb{P}}\left(\varepsilon_n^{\frac{1}{3-2H}}\right). \tag{5.34}$$

Proof. We simplify notations by omitting t in (5.32). Let and $c > 0$ and set $\gamma^* = 1/(3-2H)$. By virtue of Lemma 6.0.14, we derive from (5.26) the following result

$$\left|\frac{\widehat{Q}_{1,\varepsilon_n}}{\phi_1 b(x^1)} - 1\right| = O_{\mathbb{P}}\left(\varepsilon_n^{\gamma^*}\right).$$

Hence,

$$\begin{aligned}
\sup_n \mathbb{P}\left(\varepsilon_n^{-\gamma^*}\left|\widehat{\phi}_{1,n} - \phi_1\right| > c\right) &\leq \sup_n \mathbb{P}\left(\frac{1}{2}\phi_1^2 + \frac{1}{2}\varepsilon_n^{-2\gamma^*}\left|\frac{\widehat{Q}_{1,\varepsilon_n}}{\phi_1 b(x^1)} - 1\right|^2 > c\right) \\
&\leq \mathbb{P}\left(\phi_1^2 > c\right) + \sup_n \mathbb{P}\left(\varepsilon_n^{-2\gamma^*}\left|\frac{\widehat{Q}_{1,\varepsilon_n}}{\phi_1 b(x^1)} - 1\right|^2 > c\right) \\
&\leq \frac{\mathbb{E}\phi_1^2}{c} + \sup_n \mathbb{P}\left(\left|\frac{\widehat{Q}_{1,\varepsilon_n}}{\phi_1 b(x^1)} - 1\right| > \sqrt{c}\,\varepsilon_n^{\gamma^*}\right) \\
&\longrightarrow 0 \text{ as } c \longrightarrow \infty.
\end{aligned}$$

□

It is well known that standard kernel density estimators for the unknown density f of ϕ_i are given by

$$\widetilde{f}_{h_\varepsilon}(x) = \frac{1}{Nh_\varepsilon}\sum_{i=1}^{N} G\left(\frac{x-\phi_i}{h_\varepsilon}\right), \quad h_\varepsilon > 0,$$

where G is an integrable kernel that has to satisfy some regularity conditions. Since the random effects ϕ_i are not observed; it is natural to replace them by their estimates to obtain the following kernel density estimators

$$\begin{aligned}\widehat{f}^{(1)}_{h_\varepsilon}(x) &= \frac{1}{Nh_\varepsilon}\sum_{i=1}^{N} G\left(\frac{x-\widehat{\phi}^{(1)}_{i,n}}{h_\varepsilon}\right);\\ \widehat{f}^{(2)}_{h_\varepsilon}(x) &= \frac{1}{Nh_\varepsilon}\sum_{i=1}^{N} G\left(\frac{x-\widehat{\phi}^{(2)}_{i,n}}{h_\varepsilon}\right),\end{aligned} \tag{5.35}$$

and prove their consistency as n and N become large. We focus on the estimators $\widehat{f}^{(1)}_{h_\varepsilon}$ because the same procedures can be applied to get the asymptotic behaviour of $\widehat{f}^{(2)}_{h_\varepsilon}$.

5.4.3 Integrated squared risk bound of the estimator $\widehat{f}^{(1)}_{h_\varepsilon}$

We make the following assumptions which will be used to obtain an integrated squared risk bound

A$_3$: The kernel G is of order $l \geq 1$, that is,

$$\int_{\mathbb{R}} G(u)du = 1, \quad \int_{\mathbb{R}} u^j G(u)du = 0, \quad \text{for } j = 1, \cdots, l.$$

For the construction of such a kernel we refer to [110, p. 10]. Assume also that G is differentiable with

$$\|G\|_{L^2(\mathbb{R})} + \|G'\|_{L^2(\mathbb{R})} < \infty \quad \text{and} \quad \int_{\mathbb{R}} |u|^{\delta} |G(u)| du < \infty.$$

A$_4$: The density function f belongs to the Nikol'ski class $\mathcal{N}(\delta, R)$ defined as the set of functions $f : \mathbb{R} \longrightarrow \mathbb{R}$ whose derivatives $f^{(l)}$ of order $l = \lfloor \delta \rfloor$ exist and satisfy

$$\left[\int_{\mathbb{R}} \left(f^{(l)}(x+t) - f^{(l)}(x)\right)^2 dx\right]^{1/2} \leq R\,|t|^{\delta - l}, \text{ for all } t \in \mathbb{R},$$

where $\lfloor \phi \rfloor$ denotes the greatest integer strictly less than the real number δ.

Theorem 5.4.4.
Let the assumptions $\boldsymbol{A}_3$ and $\boldsymbol{A}_4$ be fulfilled. Assume further that there exists $\gamma > 1$ such that

$$\lim_{v\to\infty} v^{\gamma}\psi(v) < \infty, \tag{5.36}$$

where $\psi(v) := \sup_n \mathbb{P}\left((\widehat{\phi}^{(1)}_{1,n} - \phi_1)^2 > v\delta_n^2\right)$ *and* $\delta_n = \varepsilon_n^{1/(3-2H)}$. *Then, by considering estimator* $\widehat{f}^{(1)}_{h_\varepsilon}$ *given by (5.35), we have*

$$\mathbb{E}\left\|f - \widehat{f}^{(1)}_{h_\varepsilon}\right\|^2_{L^2(\mathbb{R})} \le C\left(h_\varepsilon^{2\delta} + \frac{\|G\|^2_{L^2(\mathbb{R})}}{Nh_\varepsilon} + \frac{\|G'\|^2_{L^2(\mathbb{R})}}{h_\varepsilon^3}\delta_n^2\right),$$

where $C > 0$ *is a nonnegative constant.*

Proof. By virtue of Lemmas 6.0.15 and 6.0.16, we obtain

$$\begin{aligned}\mathbb{E}\left\|\widehat{f}^{(1)}_{h_\varepsilon} - f\right\|^2_{L^2(\mathbb{R})} &= \left\|f - \mathbb{E}\widehat{f}^{(1)}_{h_\varepsilon}\right\|^2_{L^2(\mathbb{R})} + \mathbb{E}\left\|\widehat{f}^{(1)}_{h_\varepsilon} - \mathbb{E}\widehat{f}^{(1)}_{h_\varepsilon}\right\|^2_{L^2(\mathbb{R})} \\ &\le 2\left\|f - f_{h_\varepsilon}\right\|^2_{L^2(\mathbb{R})} + 2\left\|f_{h_\varepsilon} - \mathbb{E}\widehat{f}^{(1)}_{h_\varepsilon}\right\|^2_{L^2(\mathbb{R})} + \mathbb{E}\left\|\widehat{f}^{(1)}_{h_\varepsilon} - \mathbb{E}\widehat{f}^{(1)}_{h_\varepsilon}\right\|^2_{L^2(\mathbb{R})} \\ &\le 2C(l,\delta,R)h_\varepsilon^{2\delta} + \frac{\|G\|^2_{L^2(\mathbb{R})}}{Nh_\varepsilon} + 2\left\|f_{h_\varepsilon} - \mathbb{E}\widehat{f}^{(1)}_{h_\varepsilon}\right\|_{L^2(\mathbb{R})},\end{aligned} \tag{5.37}$$

where $C(l,\delta,R)$ is a nonnegative constant due to Lemma 6.0.15.

There remains to find an upper bound of the last term in (5.37). First, note that $f_{h_\varepsilon}(x) = \int_{\mathbb{R}} f(y)G_{h_\varepsilon}(x-y)dy = \mathbb{E}\left(G_{h_\varepsilon}(x-\phi_1)\right)$. Taylor's theorem with integral remainder yields

$$G_{h_\varepsilon}(x - \widehat{\phi}^{(1)}_{1,n}) - G_{h_\varepsilon}(x-\phi_1) = \frac{(\phi_1 - \widehat{\phi}^{(1)}_{1,n})}{h_\varepsilon^2}\int_0^1 G'\left(\frac{1}{h_\varepsilon}(x - \phi_1 + u(\phi_1 - \widehat{\phi}^{(1)}_{1,n}))\right)du.$$

Now, by setting $g(x,u) = G'\left(\frac{1}{h_\varepsilon}(x - \phi_1 + u(\phi_1 - \widehat{\phi}^{(1)}_{1,n}))\right)$, we have

$$\begin{aligned}\left\|f_{h_\varepsilon} - \mathbb{E}(\widehat{f}^{(1)}_{h_\varepsilon})\right\|^2_{L^2(\mathbb{R})} &= \int_{\mathbb{R}}\left[\mathbb{E}\left(G_{h_\varepsilon}(x - \widehat{\phi}^{(1)}_{1,n}) - G_{h_\varepsilon}(x-\phi_1)\right)\right]^2 dx \\ &\le \int_{\mathbb{R}}\mathbb{E}\left(G_{h_\varepsilon}(x - \widehat{\phi}^{(1)}_{1,n}) - G_{h_\varepsilon}(x-\phi_1)\right)^2 dx \\ &\le \mathbb{E}\left[\frac{(\phi_1 - \widehat{\phi}^{(1)}_{1,n})^2}{h_\varepsilon{}^4}\int_{\mathbb{R}}\left(\int_0^1 g(x,u)du\right)^2 dx\right] \\ &\le \mathbb{E}\left[\frac{(\phi_1 - \widehat{\phi}_{1,n})^2}{h_\varepsilon{}^4}\left(\int_0^1\left(\int_{\mathbb{R}} g^2(x,u)dx\right)^{1/2} du\right)^2\right].\end{aligned}$$

The last inequality given above is justified by the generalized Minkowski inequality (see [110, Lemma A.1]). By change of variables $y = \frac{1}{h_\varepsilon}\left(x - \phi_1 + u(\phi_1 - \widehat{\phi}^{(1)}_{1,n})\right)$, we get

$\int_{\mathbb{R}} g^2(x,u)dx = \|G'\|^2_{L^2(\mathbb{R})} h_\varepsilon$. Thus,

$$\left\| f_{h_\varepsilon} - \mathbb{E}(\widehat{f}^{(1)}_{h_\varepsilon}) \right\|^2_{L^2(\mathbb{R})} \leq \frac{\|G'\|^2_{L^2(\mathbb{R})}}{{h_\varepsilon}^3} \mathbb{E}(\phi_1 - \widehat{\phi}^{(1)}_{1,n})^2. \tag{5.38}$$

But we know that

$$\begin{aligned} \mathbb{E}(\phi_1 - \widehat{\phi}^{(1)}_{1,n})^2 &= \int_0^\infty \mathbb{P}\left((\phi_1 - \widehat{\phi}^{(1)}_{1,n})^2 > y\right) dy \\ &= \delta_n^2 \int_0^\infty \mathbb{P}\left((\phi_1 - \widehat{\phi}^{(1)}_{1,n})^2 > v\delta_n^2\right) dy \\ &\leq \delta_n^2 \left(1 + \int_1^\infty \psi(v)dv\right). \end{aligned} \tag{5.39}$$

Therefore, the combination of (5.37) and (5.38) with (5.39) yields the desired result. □

Corollary 5.4.5.
Set $h_\varepsilon = \delta_n^{\frac{2}{3+2\delta}}$ *and* $N = \lfloor \delta_n^{-\frac{2+4\delta}{3+2\delta}} \rfloor + 1$, *where* $\delta_n = \varepsilon_n^{1/(3-2H)}$, *we have*

$$\mathbb{E}\left\| f - \widehat{f}^{(1)}_{h_\varepsilon} \right\|_{L^2(\mathbb{R})} = O\left(\delta_n^{\frac{2\delta}{3+2\delta}}\right).$$

The integrated squared risk bound of $\widehat{f}^{(1)}_{h_\varepsilon}$ is strongly related to the convergence of the estimators $\widehat{\phi}^{(1)}_{i,n}$ in L^2-sense, which is equivalent to the uniform integrability of the sequence $\left\{ \left|\widehat{\phi}^{(1)}_{i,n}\right|^2 \ : \ n \geq 1 \right\}$, since the convergence in probability was established. Even this uniform integrability does not help us to get a precise rate of convergence of $\widehat{\phi}^{(1)}_{i,n}$ toward ϕ_i, it is not clear how to verify it under the proposed assumptions. Thus, the pointwise risk of $\widehat{f}^{(1)}_{h_\varepsilon}$ is worth being examined in the next subsection. For this end we do not need any assumption like (5.36).

5.4.4 Pointwise risk bound of the estimator $\widehat{f}^{(1)}_{h_\varepsilon}$

In this subsection, in stead of considering $\mathbf{A}_3$ and $\mathbf{A}_4$, we make the following assumptions :

A'$_3$: The density function f satisfies the Hölder condition, that is,

$$|f(u) - f(v)| \leq D\,|u - v|^\gamma, \quad \forall u, v \in \mathbb{R},$$

where D and γ are some nonnegative constants.

A'$_4$: The kernel G is differentiable and satisfies $M := \sup_u |G(u)| < \infty$, $M' := \sup_u |G'(u)| < \infty$ and $\int_{\mathbb{R}} |u|^{\gamma} |G(u)| \, du < \infty$, where γ is nonnegative constant appearing in **A'**$_3$.

Theorem 5.4.6.
Let the assumptions ***A'***$_3$ *and* ***A'***$_4$ *be fulfilled. Then*

$$\left| \widehat{f}^{(1)}_{h_\varepsilon}(x) - f(x) \right| = O_{\mathbb{P}} \left(\delta_n h_\varepsilon^{-2} \vee h_\varepsilon^{\gamma} \vee h_\varepsilon^{-1}/\sqrt{N} \right), \quad \forall x \in \mathbb{R},$$

where $\delta_n = \varepsilon_n^{1/(3-2H)}$.

Proof. Let $x \in \mathbb{R}$ and $c > 0$, for a nonnegative sequence $(\lambda_n)_{n \geq 1}$ with $\lambda_n \to 0$ (to be specified later), we have

$$\begin{aligned}
\sup_n \mathbb{P} \left(\left| \widehat{f}^{(1)}_{h_\varepsilon}(x) - f(x) \right| > c\lambda_n \right) &\leq \sup_n \mathbb{P} \left(\left| \widehat{f}^{(1)}_{h_\varepsilon}(x) - \widetilde{f}_{h_\varepsilon}(x) \right| > \frac{c}{3}\lambda_n \right) \\
&\quad + \sup_n \mathbb{P} \left(\left| \widetilde{f}^{(1)}_{h_\varepsilon}(x) - \mathbb{E}\widetilde{f}^{(1)}_{h_\varepsilon}(x) \right| > \frac{c}{3}\lambda_n \right) \\
&\quad + \sup_n \mathbb{P} \left(\left| \mathbb{E}\widetilde{f}^{(1)}_{h_\varepsilon} - f(x) \right| > \frac{c}{3}\lambda_n \right) \\
&\leq \sup_n \mathbb{P} \left(\sup_x \left| \widehat{f}^{(1)}_{h_\varepsilon}(x) - \widetilde{f}_{h_\varepsilon}(x) \right| > \frac{c}{3}\lambda_n \right) \\
&\quad + \frac{9}{c^2\lambda_n^2} \sup_n \mathbb{E} \left| \widetilde{f}^{(1)}_{h_\varepsilon}(x) - \mathbb{E}\widetilde{f}^{(1)}_{h_\varepsilon}(x) \right|^2 \\
&\quad + \sup_n \mathbb{P} \left(\sup_x \left| \mathbb{E}\widetilde{f}^{(1)}_{h_\varepsilon}(x) - f(x) \right| > \frac{c}{3}\lambda_n \right). \qquad (5.40)
\end{aligned}$$

By virtue of Lemmas 6.0.17-6.0.20, the RHS in (5.40) tends to zero as $c \to \infty$, provided that $\lambda_n \geq \left(\delta_n h_\varepsilon^{-2} \vee h_\varepsilon^{\gamma} \vee h_\varepsilon^{-1}/\sqrt{N} \right)$. □

Corollary 5.4.7.
Set $h_\varepsilon = \delta_n^{\frac{1}{2+\gamma}}$ *and* $N = \lfloor \delta_n^{-\frac{2+2\gamma}{2+\gamma}} \rfloor + 1$, *where* $\delta_n = \varepsilon_n^{1/(3-2H)}$. *We have*

$$\left| \widehat{f}^{(1)}_{h_\varepsilon}(x) - f(x) \right| = O_{\mathbb{P}} \left(\delta_n^{\frac{\gamma}{2+\gamma}} \right), \quad \forall x \in \mathbb{R}.$$

5.4.5 Implementation issues and numerical applications

In this section, we consider SDE's given by (5.20) with ϕ_i are either Gaussian or Beta random effects. In this case, all assumptions made on the associated density function are met. For the sake of optimizing our programs, we choose $K(u)$ to be rectangular kernel given by $K(u) = \frac{1}{2}\chi_{(|u| \leq 1)}$. This kernel is very good for random effects taking values in

$\mathbb{R}$ (Gaussian). But, in case where random effects $\phi_i \in \mathbb{K}$ (compact set), one has to trim the proposed estimators by $\chi_{\mathbb{K}}$. The good choice of the kernel $G(u)$ would be to consider the Gaussian kernel, because it satisfies the assumptions A_3 and A_4'. Two models are considered : $b(x) = 1$ and $b(x) = \dfrac{x}{\sqrt{1+x^2}}$ with curves illustrated, respectively, by Figure 5.8 and Figure 5.9 below. Results, are satisfactory overall. Increasing N and/or decreasing the value of ε improves the accuracy of both estimates $\hat{f}_{h_\varepsilon}^{(1)}$ and $\hat{f}_{h_\varepsilon}^{(2)}$.

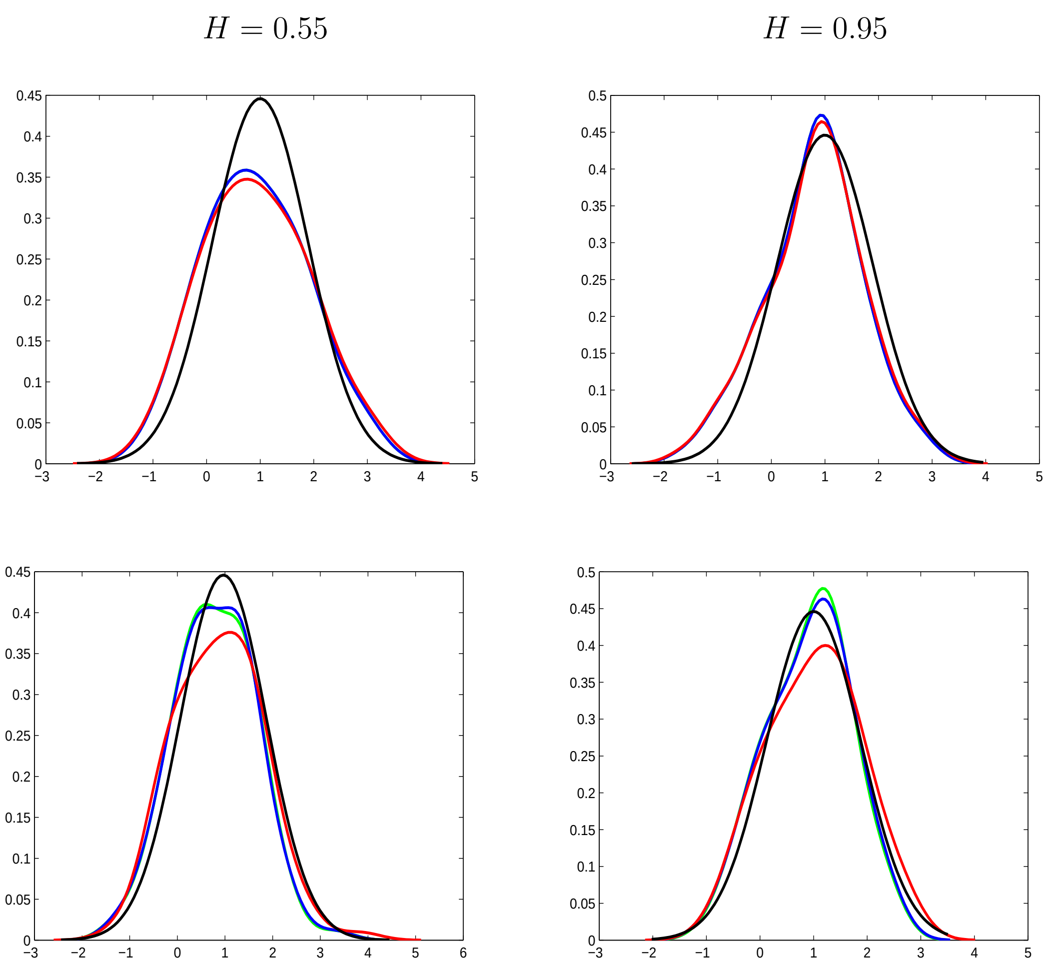

Figure 5.8. Kernel estimates $\hat{f}^{(1)}_{h_\varepsilon}$ and $\hat{f}^{(2)}_{h_\varepsilon}$ of f. The exact curve of f and its estimate based on true values of random effects are given in green, blue, black and red, respectively. The first row illustrates the case where $b(x) = 1$ and the second row illustrates the case where $b(x) = \dfrac{x}{\sqrt{1+x^2}}$, with Gaussian random effects $\phi_i \sim \mathcal{N}(1, 0.8)$ and $(T, N, \varepsilon) = (5, 10^2, 10^{-2})$. (For interpretation of the references to colour the reader is referred to the electronic version of this thesis).

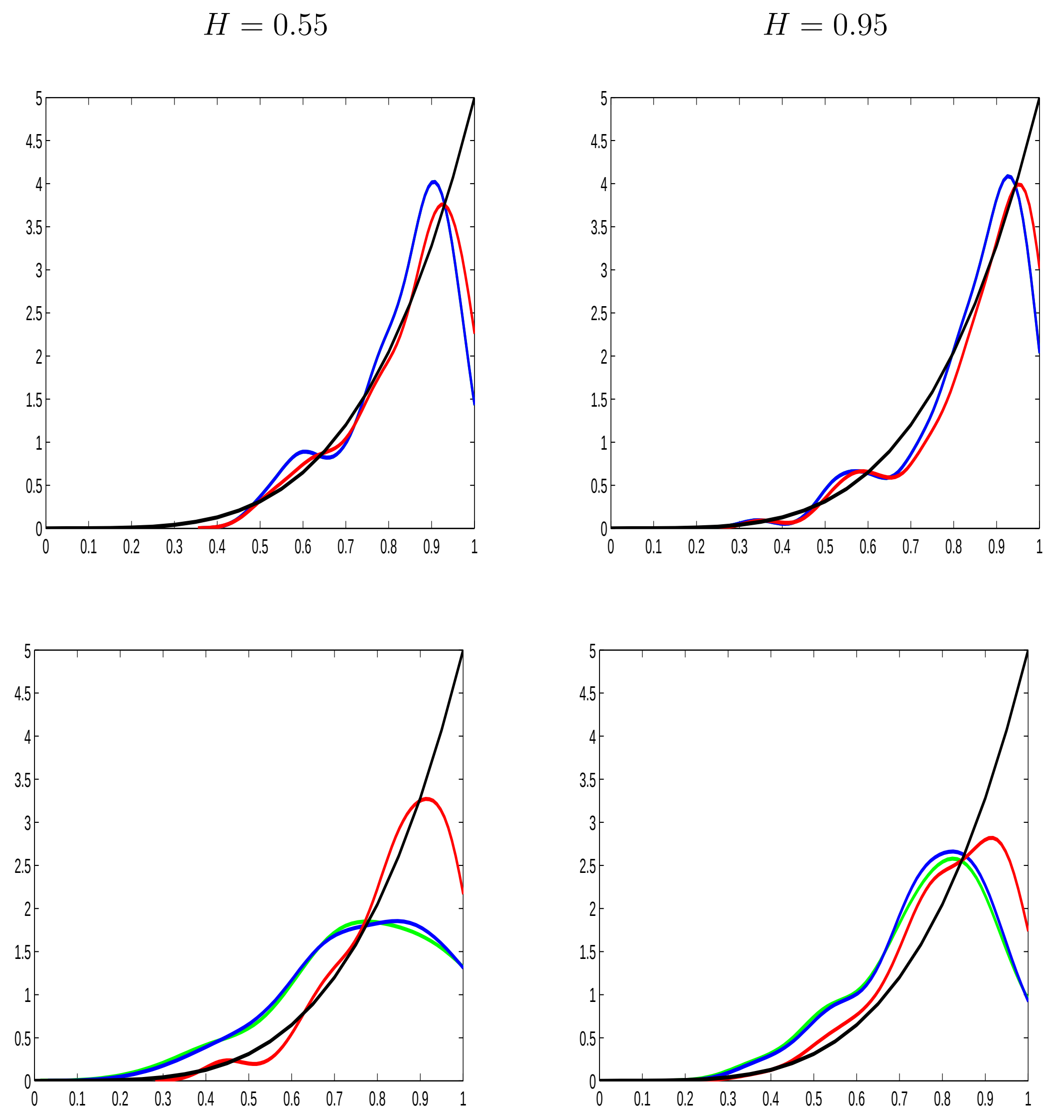

Figure 5.9. Kernel estimates $\hat{f}^{(1)}_{h_\varepsilon}$ and $\hat{f}^{(2)}_{h_\varepsilon}$ of f. The exact curve of f and its estimate based on true values of random effects are given in green, blue, black and red, respectively. The first row illustrates the case where $b(x) = 1$ and the second row illustrates the case where $b(x) = \dfrac{x}{\sqrt{1+x^2}}$, with Beta random effects $\phi_i \sim \mathcal{B}(5, 1)$ and $(T, N, \varepsilon) = (5, 10^2, 10^{-2})$. (For interpretation of the references to colour the reader is referred to the electronic version of this thesis).

CHAPTER 6
CONCLUDING REMARKS AND PERSPECTIVES

The intention of this thesis was to provide statistical methods for performing both parametric and non parametric estimation on two classes of REM with dynamics ruled by

1. Nonlinear SDE's with generalized random effects;

2. Linear FSDE's with random effects, when the Hurst index $H > 1/2$.

In Chapter 4, we considered the first class of REM with dynamics ruled by (4.1). As starting point, a class of estimators of the population parameters has been proposed in Section 4.1. We note that SDE models incorporating random effects have been considered in few recent works (e.g., [25]) focused on models with linear drift. Other papers (see, [83, 82, 81]) considered a nonlinear models and provide only the consistency results. Our proposed estimators concern generalized linear and nonlinear drift, and are shown to be consistent and asymtotically normal. The conditions we provided are not necessary as general as those conditions in the literature. These conditions can be reduced, however, the general conditions may be very difficult to verify with mathematical rigorousness. Simulation results for the example 4.1.5 have shown that our proposed estimators are close to the true parameter values. Compared to Delattre's results (Table 4 in [25]), our results computed from 100 datasets are excellent, even both N and R are not too large. Results obtained for nonlinear models (Example 4.1.2) are very satisfactory, and the accuracy can be improved by increasing only N. Compared to N (number of subjects), increasing R have no significant impact on the properties of estimators and consume more time. To weaken the imposed assumptions on our model, we expanded the likelihood function in Section 4.2 by means of iterated Itô integrals (with random integrands). To implement examples for which the weakned assumptions are met, one has to overcome the problem of simulating the multiple integrals of the form

$$I_T(h) := \int_0^T \int_0^{t_1} \cdots \int_0^{t_{n-1}} h(t_1, t_2, \cdots, t_n, \omega) dW_{t_n} \cdots dW_{t_2} dW_{t_1}, \ n \geq 2. \tag{6.1}$$

This problem is only solved in literature for a particular case where $h(\cdot)$ is a tensor power of a nonrandom function in $L^2([0,T])$ (e.g., [28]). Simulating the integral (6.1) in the classical way, recursively, presents a computational chanllenge. In Chapter 5, we considered the second class of REM (5.1), for which three cases are separately discussed in sections within this chapter

- Case 1: $b(t,x) = 1$ and $\sigma(t,x) = 1$
- Case 2: $b(t,x) = b(t)$ and $\sigma(t,x) = \sigma(t)$
- Case 3: $a(x) = 0$, $b(t,x) = b(x)$ and $\sigma(t,x) = \varepsilon$

In Section 5.2, we considered Case 1, for which we have provided a fully Likelihood parametric estimation. We are essentially concerned with the estimation of Hurst index, as well as with the mean and variance estimators of the random effect that has a Gaussian distribution. All qualitative and asymptotic properties of the estimators are obtained, when the population of subjects becomes large. The study of this case suggests several important directions for future research. First, what are the asymptotic properties of the Maximum Likelihood estimators for μ and ω^2 when the Hurst index H is unknown? Given that the model is fully parameterized, one may wish to estimate H, μ and ω^2 simultaneously. Second, the study of this case assumes that the model is linear and the diffusion is constant and equals 1. This assumption is not verified in almost all real applications. So, one can use, for example, Euler schemes approximation. However, it is not clear how to get an explicit approximation for the Maximum Likelihood function. Such extension would be worth being studied from both theoretical and application points of view. Third, as mentioned in Section 5.2, we may estimate the population parameters by using double asymptotic framework. Such an idea was applied in Section 5.3 to tackle the non-parametric estimation for REM of type: Case 2. In this section, we addressed the open research question of how to estimate the density of random effects in FSDE's in a non-parametric fashion. To that end, we considered N i.i.d processes $X^i(t),\ 0 \leq t \leq T,\ i = 1, \cdots, N$, where the dynamics of X^i is described by an FSDE including a random effect ϕ_i . The non-parametric estimation of the density of ϕ_i was investigated for a general linear model of the form $dX_t = (a(X_t) + \phi b(t))dt + \sigma(t)dW_t^H$, where $b(\cdot)$ and $\sigma(\cdot)$ are known functions, but $a(\cdot)$ is possibly unknown. We studied the asymptotic behaviour of the proposed density estimators for the whole range $H \in (1/2, 1)$, built kernel density estimators and studied their L^2-risk as both N and T tended to infinity. We also provided histogram estimators in a specific case where f has compact support, which was for two reasons: First, we aimed to simplify technical computations, and cases where the random effects density f has unbounded support are less important, since data could always be mapped monotonically to $[0,1]$. Second, densities with unbounded support occur less often in practice. For the proposed histogram estimators, we provided their L^1 -risk for both N and $T = T(N)$ tending to infinity.

With Mémin's results (see, Theorem 2.2.2) , we can have only a lower inequality of the form

$$\mathbb{E}\left|\int_0^T h(t)dW_t^H\right|^p \geq C_{H,p}\,\|h\|^p_{L^{1/H}([0,T])}\,,\ \text{when } H<1/2.$$

This inequality would not help us to prove the convergence of both $\widehat{\phi}_{i,T}$ and $\widetilde{\phi}_{i,T}$ toward ϕ_i, when $H < 1/2$. This suggests to apply more advanced techniques, such as Malliavin calculus. As mentioned before, our results are very good even for the case $H < 1/2$, which is not supported by our theoretical framework. This is can be justified by recent results of Hu et al., [54], where the authors provided moment estimates and maximal inequality for divergence integrals w.r.t fBm, when $H \in (0, 1/2) \cup (1/2, 1)$. It is then convenient to see our integral of interest $\int_0^T \frac{b(t)}{\sigma(t)}dW_t^H$ as divergence integral (or Skorohod integral), so that for $H < 1/2$ and $p \geq 2$, we have

$$\mathbb{E}\left|\int_0^T \frac{b(t)}{\sigma(t)}dW_t^H\right|^p \leq CT^{pH}(1 + T^{p\beta} + T^{p\beta+p\lambda}), \tag{6.2}$$

where $C > 0$ is constant independent of T, $\lambda \in (0, H]$ and $\beta > 1/2 - H$, provided that

$\mathbf{U}_3$: **i)** $\sup_{t\geq t}\left|\frac{b(t)}{\sigma(t)}\right| < \infty$

ii) $\left|\frac{b(t)}{\sigma(t)} - \frac{b(s)}{\sigma(s)}\right| \leq K(t-s)^\beta,$ for all $t > s \geq 0$

Our example ($b(t) = \sigma = 1$) implemented in Subsection 5.3.3 satisfies $\mathbf{A}_3$, with any arbitrary β. Choosing λ, β such that $1/2-H < \beta \leq 1/2-\lambda$, will guarantee the convergence of $\widehat{\phi}_{i,T}$ and $\widetilde{\phi}_{i,T}$ toward ϕ_i, when $H < 1/2$. An interesting extension of this study would be to consider models with nonlinear drift. In this case, one has to face the problem of estimating random effects ϕ_i . Methods of parametric estimation, such as the maximum likelihood technique, may help to estimate these random effects. The essential idea of non-parametric estimation within this chapter is to proceed in two steps :

- Estimate random effects;
- Use classical density estimators with random effects replaced by their estimates.

The third case of our REM was treated in Section 5.4. Both L^2-risk and pointwise risk (in probability) of our estimators are examined. The simulations were performed for different drift term and different density of random effects (Gaussian and Beta distributions). The inference of short-range models with random effects persists as challenge, and is worth being studied from both theoretical and application points of view.

REFERENCES

[1] Ait-Sahalia, Y. *Closed-form likelihood expansions for multivariate diffusions.* The Annals of Statistics (2008),36(2), 906-937.

[2] Ait-Sahalia, Y. *Maximum likelihood estimation of discretely sampled diffusions: a closed-form approximation approach.* Econometrica (2002),70(1), 223-262.

[3] Allen, L. J. S. *An introduction to stochastic epidemic models.* Mathematical Epidemiology, Vol. 1945 of Lecture Notes in Mathematics, Springer Berlin Heidelberg, (2008),81-130.

[4] Alonso, D., McKane, A. J., Pascual, M. *Stochastic amplification in epidemics.* Journal of the Royal Society Interface (2007),4(14), 575-582.

[5] Alos, E., Mazet, O., Nualart, D. *Stochastic calculus with respect to Gaussian processes.* Ann. Probab. (2001), 29, 766-801.

[6] Antic, J., Laffont, C.M., Chafai, D., Condorcet, D. *Comparison of nonparametric methods in nonlinear mixed effects models*, Comput. Statist. Data Anal,(2009), 53, 642-656.

[7] Bachelier, L. *Theory of Speculation, the Random Character of Stock Prices, translation of Bacheliers 1900 doctoral thesis.* Cambridge: MIT.(1964).

[8] Barbour, A. D. *On a functional central limit theorem for Markov population processes.* Advances in Applied Probability (1974),6(1), 21-39.

[9] Beskos, A., Kalogeropoulos, K., Pazos, E. *Advanced MCMC methods for sampling on diffusion pathspace.* Stochastic Processes and their Applications (2013),123(4), 1415-1453.

[10] Beskos, A., Papaspiliopoulos, O., Roberts, G. O. *Monte Carlo maximum likelihood estimation for discretely observed diffusion processes.* Annals of Statistics. (2009),37, 223-245.

[11] Beskos, A., Papaspiliopoulos, O., Roberts, G. O., Fearnhead, P. *Exact and computationally efficient likelihood-based estimation for discretely observed diffusion processes (with discussion).* Journal of the Royal Statistical Society: Series B (Statistical Methodology). (2006),68, 333-382.

[12] Bibby, B. M., Sørensen, M. *Simplified estimating functions for diffusion models with a high-dimensional parameter.* Scandinavian Journal of Statistics (2001), 28(1), 99-112.

[13] Bishwal, J. P.N. *Parameter estimation in stochastic differential equations.* Berlin Heidelberg: Springer-Verlag, (2008).

[14] Black, F., Scholes, M. *The pricing of options and corporate liabilities.* Journal of Political Economy (1973),81(3), 637-654.

[15] Brown, R. *A brief account of microscopical observations made in the months of june, july and august 1827, on the particles contained in the pollen of plants; and on the general existence of active molecules in organic and inorganic bodies.*Philosophical Magazine Series2 (1828),4(21), 161-173.

[16] Capasso, V., Morale, D. *Stochastic modelling of tumour-induced angiogenesis.* Journal of Mathematical Biology (2009),58(1-2), 219-233.

[17] Chen, W. Y., Bokka, S. *Stochastic modeling of nonlinear epidemiology.* Journal of Theoretical Biology (2005),234(4), 455-470.

[18] Chiarella, C., Hung, H., Tô, T. D. *The volatility structure of the fixed income market under the HJM framework: a nonlinear filtering approach.* Computational Statistics & Data Analysis (2009),53(6), 2075-2088.

[19] Coeurjolly, J. F. *Estimating the parameters of a fractional Brownian motion by dis-*

crete variations of its sample paths. Stat Inference Stoch Process,(2001), 4: 199-227.
[20] Cohen, S., Istas, J. *Fractional fields and applications.* Springer-Verlag, Berlin Heidelberg, (2013).
[21] Collins, J. J., De Luca, C.J. *Upright, correlated random walks: A statistical-biomechanics approachto the human postural control system*, Chaos, (1995), 5, 1, 57-63.
[22] Comte, F., Samson, A. *Nonparametric estimation of random effects densities in linear mixed-effects model*, J. Nonparametr. Stat,(2012), 24 (4) 951-975.
[23] Cox, J. C., Ingersoll, Jr. J. E., Ross, S. A. *An intertemporal general equilibrium model of asset prices.* Econometrica (1985),53(2), 363-384.
[24] Dargatz, C. *A diffusion approximation for an epidemic model.* Discussion paper, Sonderforschungsbereich 386 der Ludwig-Maximilians-Universitt Mnchen 517. (2007).
[25] Delattre, M., Genon-Catalot, V., Samson, A. *Maximum likelihood estimation for stochastic differential equations with random effects.* Scand. J. Stat. (2012),**40**, 322-343.
[26] Dellacherie, C., Meyer, P. A. *Probabilities and Potential.* North-Holland, Amsterdam. (1982).
[27] Devroye, L., Gyorfi, L. *Nonparametric density estimation: the L_1 view.* NewYork:JohnWiley and Sons Inc. (1985).
[28] Di Nunno, G., Øksendal, B., Proske, F. *Malliavin calculus for lévy processes with applications to finance.* Heidelberg: Springer-Verlag, (2009).
[29] Ditlevsen, S., De Gaetano, A. *Mixed effects in stochastic differential equation models.* REVSTAT – Statist. J,(2005a), 3, 137-153.
[30] Donnet, S., Foulley, J.-L., Samson, A. *Bayesian analysis of growth curves using mixed models defined by stochastic differential equations.* Biometrics, (2010),66(3):733-741.
[31] Doukhan, P., Oppenheim, G., Taqqu, M.S. *Theory and Applications of Long Range-Dependence.* Birkhauser, (2003).
[32] Durham, G. B., Gallant, A. R. *Numerical techniques for maximum likelihood estimation of continuous-time diffusion processes.* Journal of Business and Economic Statistics (2002),20, 297-338.
[33] Durrett, R. *Stochastic calculus: A practical introduction.* CRC Press, Inc. (1996).
[34] Einstein, A. *The theory of Brownian movement.* Annalen der Physik 17, 549. (1905).
[35] Elerian, O., Chib, S., Shephard, N. *Likelihood inference for discretely observed nonlinear diffusions.* Econometrica (2001),69, 959-993.
[36] El Omari, M., El Maroufy, H. *Nonparametric estimation for small fractional diffusion processes with random effects*, Stochastic Analysis and Applications, (2020). DOI: 10.1080/07362994.2020.1765801
[37] El Omari, M., El Maroufy, H. Fuchs, C. *Statistical inference for fractional diffusion process with random effects at discrete observations*, (2019b). arXiv preprint arXiv: 1912.01463
[38] El Omari, M., El Maroufy, H. Fuchs, C. *Non parametric estimation for fractional diffusion processes with random effects*, *Statistics.* (2019a). DOI: 10.1080/02331888.2019.1581783
[39] Eraker, B. *MCMC analysis of diffusion models with application to finance.* Journal of Business and Economic Statistics (2001),19(2), 177-191.
[40] Fearnhead, P., Giagos, V., Sherlock, C. *Inference for reaction networks using the linear noise approximation.* Biometrics (2014),70(2), 457-466.
[41] Fearnhead, P. *The stationary distribution of allele frequencies when selection acts at unlinked loci.* Theoretical Population Biology (2006),70(3), 376-386.
[42] Fernandes, A., Atchley, W. *Gaussian quadrature formulae for arbitrary positive mea-*

sures. Evol. Bioinform. Online. (2006),**2**, 261-269.

[43] Finkenstdt, B., Heron, E. A., Komorowski, M., Edwards, K., Tang, S., Harper, C. V., Davis, J. R. E., White, M. R. H., Millar, A. J., Rand, D. A. *Reconstruction of transcriptional dynamics from gene reporter data using differential equations.* Bioinformatics (2008),24(24), 2901-2907.

[44] Fogelson, A. *A mathematical model and numerical method for studying platelet adhesion and aggregation during blood clotting.* Journal of Computational Physics (1984),56(1), 111-134.

[45] Fuchs, C. *Inference for Diffusion Processes with Applications in Life Sciences.* Springer, Heidelberg.(2013)

[46] Genon-Catalot, V., Laredo, C. *Stationary distributions for stochastic differential equations with random effects and statistical applications.* MAP5 -09. (2013), ¡hal-00807258v1¿

[47] Gilioli, G., Pasquali, S., Ruggeri, F. *Bayesian inference for functional response in a stochastic predatorprey system.* Bulletin of Mathematical Biology (2008),70(2), 358-381.

[48] Golightly, A., Henderson, D. A., Sherlock, C. *Delayed acceptance particle MCMC for exact inference in stochastic kinetic models* Statistics and Computing (2015),25(5), 1039-1055.

[49] Golightly, A., Wilkinson, D. J. *Bayesian inference for nonlinear multivariate diffusion models observed with error.* Computational Statistics & Data Analysis (2008),52(3), 1674-1693.

[50] Granger, C. W. J. *The typical spectral shape of an economic variable*, Econometrica,(1966), 34 , 150-161.

[51] Gray, A., Greenhalgh, D., Hu, L., Mao, X., Pan, J. *A stochastic differential equation SIS epidemic model.* SIAM Journal on Applied Mathematics (2011),71(3), 876-902.

[52] Gripenberg, G., Norros, I. *On the prediction of fractional Brownian motion.* J. Appl. Probab. (1996), 33, 400-410.

[53] Heydari, J. J., Lawless, C., Lydall, D. A., Wilkinson, D. J. *Fast Bayesian parameter estimation for stochastic logistic growth models.* BioSystems (2014),122, 55-72.

[54] Hu, Y., Nualart, D., Zhou, H. *Drift parameter estimation for nonlinear stochastic differential equations driven by fractional Brownian motion.* Stochastics, (2019), 1-25.

[55] Hu, Y., Nualart, D., Xiao, W., Zhang, W. *Exact maximum likelihood estimator for drift fractional Brownian motion at discrete observation.* Acta Math. Sci. Ser. B Engl. Ed, (2011), 31(5), 1851-1859.

[56] Hu, Y., Nualart, D. *Parameter estimation for fractional Ornstein - Uhlenbeck processes.* Statist. Probab. Lett, (2010), 80(11-12), 1030-1038.

[57] Huang, S. T., Cambanis, S. *Stochastic and multiple Wiener integrals of Gaussian processes.* Ann. Probab., (1978),6, 585-614.

[58] Jarne, A., Commenges, D., Villain, L., Prague, M., Lévy, Y., Thibaut, R. *Modeling CD4 + T cells dynamics in HIV-infected patients receiving repeated cycles of exogenous interleukin 7.* Annals of Applied Statistics, (2017),11(3):1593-1616.

[59] Kalogeropoulos, K., Roberts, G. O., Dellaportas, P. *Inference for stochastic volatility models using time change transformations.* The Annals of Statistics (2010),38(2), 784-807.

[60] Karatzas, I., Shreve, S. E. *Brownian Motion and Stochastic Calculus, 2nd Edition.* Springer-Verlag New York Berlin Heidelberg.(1996).

[61] Karlin, S., Taylor, H.E. *A second course in stochastic processes*, Elsevier, (1981).

[62] Kermack, W. O., McKendrick, A. G. *A contribution to the mathematical theory of*

epidemics. Proceedings of the Royal Society of London A: Mathematical, Physical and Engineering Sciences (1927), 115(772), 700-721.

[63] Kleptsyna, M. L., Le Breton, A. *Statistical analysis of the fractional Ornstein-Uhlenbeck type process.* Stat. Inference Stoch. Process, (2002), 5(3), 229-248.

[64] Kolmogorov, A.N. *Wienersche Spiralen und einige andere interessante Kurven im Hilbertschen Raum.* C.R.(Doklady) Acad. URSS (N.S). (1940), 26, 115-118.

[65] Komorowski, M., Finkenstdt, B., Rand, D. A. *Using a single fluorescent reporter gene to infer half-life of extrinsic noise and other parameters of gene expression.* Biophysical Journal (2010),98(12), 2759-2769.

[66] Kou, S. C., Olding, B. P., Lysy, M., Liu, J. S. *A multiresolution method for parameter estimation of diffusion processes.* Journal of the American Statistical Association (2012),107, 1558-1574.

[67] Kroese, D. P., Botev, Z. I. *Spatial process generation.* (2013). available arXiv:1308.0399v1 [stat.CO]

[68] Kuklinski, W. S., Chandra, K., Ruttimann, U.E., WEBBER, R.L. *Application of Fractal Texture Analysis to Segmentation of Dental Radiographs SPIE*, Medical Imaging III: Image Processing,(1989), 1092 , 111-117.

[69] Kutoyants, Yur, A. *Statistical inference for ergodic diffusion processes.* Springer-Verlag, London. (2004).

[70] Laird, N. M., J. H. Ware. *Random-effects models for longitudinal data.* Biometrics, (1982),38: 963-974.

[71] Lavielle, M. *Mixed effects models for the population approach: models, tasks, methods and tools.* CRC Press, (2014).

[72] Leung, H. K. *Expansion of the master equation for a biomolecular selection model.* Bulletin of Mathematical Biology (1985),47(2), 231-238.

[73] Liptser, R., Shiryaev, A. *Statistics of random processes I: general theory.* Springer-Verlag, New York. (2001).

[74] Mandelbort, B.B., Van Ness Wallis, J.W. *Fractional Brownian motions, fractional noises and applications*, SIAM review, (1968), 10 , 422-437.

[75] Mcleod, A. I., Hipel, K.W. *Preservation of the Rescaled Adjusted Range: a reassessment of the Hurst exponent Water Ressources*, Research,(1978), 14, 3 , 491-508.

[76] McNeil, D. R. *Diffusion limits for congestion models.* Journal of Applied Probability (1973),10(2), 368-376.

[77] Mémin, J., Mishura, Y., Valkeila, E. *Inequalities for the moments of Wiener integrals with respect to a fractional Brownian motion.* Stat. Probab. Lett. (2001),51, 197-206.

[78] Merton, R. C. *Theory of rational option pricing.* The Bell Journal of Economics and Management Science (1973),4(1), 141-183.

[79] Mishra, M. N., Prakasa Rao, B. L. S. *Nonparametric estimation of linear multiplier for fractional diffusion processes.* Stochastic analysis and applications, (2011), 29(4), 706-712.

[80] Mishura, Y. S. *Stochastic calculus for fractional Brownian motion and related processes.* Lecture Notes in Mathematics, (2008), vol. 1929. Springer, Berlin.

[81] Nie, L.*Convergence rate of the MLE in generalized linear and nonlinear mixed-effects models: theory and applications.* J. Statist. Plann. Inference. (2007),**137**, 1787-1804.

[82] Nie, L. *Strong consistency of the maximum likelihood estimator in generalized linear and nonlinear mixed-effects models.* Metrika. (2006),**63**, 123-143.

[83] Nie, L., Yang, M.*Strong consistency of the MLE in nonlinear mixed-effects models with large cluster size.* Sankhya: The Indian J. Stat. (2005),**67**, 736-763.

[84] Nualart, D. *The Malliavin calculus and related topics.* Berlin: Springer. (2006).

[85] Nualart, D. *Stochastic integration with respect to fractional Brownian motion and applications.* Stochastic Models (Mexico City, 2002), Contemp. Math. 336 Amer. Math. Soc. Providence, RI, (2003), 3-39.
[86] Nualart, D., Rascanu, A. *Differential equations driven by fractional Brownian motion.* Collect Math, (2002), 53, 55-81.
[87] Øksendal. B. *Stochastic Differential Equations.* Universitext. sixth edition, Berlin: Springer-Verlag, (2003).
[88] Pedersen, A. R. *A new approach to maximum likelihood estimation for stochastic differential equations based on discrete observations.* Scandinavian Journal of Statistics (1995),22(1), 55-71.
[89] Picchini, U., De Gaetano, A., DITLEVSEN, S. *Stochastic differential mixed-effects models.* Scand. J. Statist,(2010), 37 , 67-90.
[90] Picchini, U., Ditlevsen, S. *Practicle estimation of a high dimensional stochastic differential mixed-effects models.* Comput. Statist. Data. Anal, (2011), 55 , 1426-1444.
[91] Pinheiro, J., Bates, D. *Mixed-effects models in S and S-PLUS.* Springer Science, New York, (2006).
[92] Prakasa Rao, B. L. S. *Nonparametric estimation of trend for stochastic differential equations driven by mixed fractional Brownian motion.* Stochastic Analysis and Applications, (2019), 37(2), 271-280.
[93] Prakasa Rao, B.L.S. *Statistical inference for fractional diffusion processes.* John Wiley & Sons.(2011).
[94] Prakasa Rao, B.L.S. *Statistical Inference for Fractional Diffusion Processes.* Wiley Series in Probability and Statistics, Wiley, Chichester. (2010).
[95] Prakasa Rao, B.L.S. *Parametric estimation for linear stochastic differential equations driven by fractional Brownian motion.* Random Oper. Stoch. Equ, (2003), 11(3), 229-242.
[96] Ramshaw, J. *Augmented langevin approach to fluctuations in nonlinear irreversible processes.* Journal of Statistical Physics (1985),38(3-4), 669-680.
[97] Ribba, B., Holford, N. H., Magni, P., Trocniz, I., Gueorguieva, I., Girard, P., Sarr, C., Elishmereni, M., Kloft, C., Friberg, L. E. *A review of mixed-effects models of tumor growth and effects of anticancer drug treatment used in population analysis.* CPT: Pharmacometrics & Systems Pharmacology, (2014)3(5):1-10.
[98] Roberts, G. O., Stramer, O. *On inference for partially observed nonlinear diffusion models using the Metropolis-Hastings algorithm.* Biometrika (2001),88(3), 603-621.
[99] Rogers, L.C.G. *Arbitrage with fractional Brownian motion.* Math. Finance 7, (1997), 95-105, A.L.
[100] Schauer, M., van der Meulen, F., van Zanten, H. *Guided proposals for simulating multi-dimensional diffusion bridges.* Bernoulli. (2016).
[101] Schervish, M. J. *Theory of Statistics.* Springer-Verlag, New York. (1995).
[102] Sermaidis, G., Papaspiliopoulos, O., Roberts, G. O., Beskos, A., Fearnhead, P. *Markov chain Monte Carlo for exact inference for diffusions.* Scandinavian Journal of Statistics (2013),40, 294-321.
[103] Sheiner, L. B, Beal, S. L., *Evaluation of methods for estimating population pharmacokinetic parameters. i. michaelis-menten model: routine clinical pharmacokinetic data.* Journal of Pharmacokinetics and Biopharmaceutics, (1980),8(6):553-571.
[104] Shiryaev, A. *Essentials of Stochastic Finance.* World Scientific, (1999).
[105] Shiryaev, A N., Boas, R. P.*Probability.* 2nd Edition ISBN 0-387-94549-0 Springer, (1996)
[106] Stramer, O., Bognar, M., Scheider, P. *Bayesian inference for discretely sampled*

Markov processes with closed-form likelihood expansions. Journal of Financial Econometrics (2010),8, 450-480.

[107] Stramer, O., Bognar, M. *Bayesian inference for irreducible diffusion processes using the pseudo-marginal approach.* Bayesian Analysis (2011),6(2), 231-258.

[108] Tian, T., Burrage, K., Burrage, P. M., Carletti, M. *Stochastic delay differential equations for genetic regulatory networks.* Journal of Computational and Applied Mathematics (2007),205(2), 696-707.

[109] Torne, C. W., Agers, H., Jonsson, E. N., Madsen, H., Nielsen, H. A. *Non-linear mixed-effects pharmacokinetic/pharmacodynamic modelling in NLME using differential equations.* Computer Methods and Programs in Biomedicine, (2004),76(1):31-40.

[110] Tsybakov, A.B. *Introduction to nonparametric estimation.* New York: Springer; (2009).

[111] Tuckwell, H. C. *Diffusion approximations to channel noise.* Journal of Theoretical Biology (1987),127(4), 427-438.

[112] Tudor, C. A., Viens, F. G. *Statistical aspects of the fractional stochastic calculus.* Ann. Statist,(2007), 35(3), 1183-1212.

[113] van der Vaart, A. W. *Asymptotic statistics*, vol. 3 of Cambridge Series in Statistical and Probabilistic Mathematics. Cambridge University Press, Cambridge. (1998). ISBN 0-521-49603-9; 0-521-78450-6.

[114] van Kampen, N. G. *The validity of nonlinear langevin equations.* Journal of Statistical Physics (1981),25(3), 431-442.

[115] Walsh, J. B. *A stochastic model of neural response.* Advances in Applied Probability (1981),13(2), 231-281.

[116] Wiener, N. *Differential-space.* Journal of Mathematics and Physics (1923), 2(1-4), 131-174.

[117] Willinger, W., Taqqu, M. S., Leland, W. E., Wilson, D. V. *Self-similarity in high-speed packet traffic: analysis and modeling of Ethernet traffic measurements.* Statistical science, (1995), 10(1), 67–85.

[118] Yaglom, A.M. *Correlation theory of processes with random stationary nth increments.* AMS Transl. 2 (1958), 8, 87–141.

[119] Young, L. C. *An inequality of the Holder type connected with Stieltjes integration.* Acta Math. 67 (1936), 251–282

APPENDICES

APPENDIX A.

Proof. **(Proof of Proposition 4.1.1)** Under $\mathbf{A}_1$-$\mathbf{A}_4$, the first part is classical result which follows from Theorem 2.1.16 (for more details see, [73, p.292]). Now, it remains to construct the exact likelihood based on the marginal densities of $\left(X^i(t),\ 0 \le t \le T_i\right)$. Let h be a nonnegative measurable function on C_{T_i}. Since $(\phi_i, X^i) \to h(X^i)$ is measurable on $\mathbb{R}^p \times C_{T_i}$, we have $\mathbb{E}_{\mathcal{Q}_\theta^i}(h(X^i)) = \mathbb{E}_{\mathbb{P}_\theta^i}(h(X^i))$, but

$$\begin{aligned}\mathbb{E}_{\mathcal{Q}_\theta^i}(h(X^i)) &= \mathbb{E}_{\mathcal{Q}_\theta^i}\left(\mathbb{E}_{\mathcal{Q}_\theta^i}(h(X^i)|\phi_i)\right)\\ &= \mathbb{E}_{\mathcal{Q}_\theta^i}(\psi(\phi_i)) = \int_{\mathbb{R}} g(\varphi,\theta)\psi(\varphi)d\nu(\varphi),\end{aligned}$$

with

$$\begin{aligned}\psi(\varphi) &= \mathbb{E}_{\mathcal{Q}_\theta^i}\left(h(X^i)|\phi_i=\varphi\right) = \int_{C_{T_i}} h(x)d\mu_{X^i_{\varphi,x^i}}(x)\\ &= \int_{C_{T_i}} h(x)\left[\frac{d\mu_{X^i_{\varphi,x^i}}}{d\mu_{X^i_{\varphi_0,x^i}}}(x)\right]d\mu_{X^i_{\varphi_0,x^i}}(x) = \mathbb{E}_{\mathbb{P}^i}\left(h(X^i)L_{T_i}(X^i,\varphi,\varphi_0)\right).\end{aligned}$$

By virtue of Fubini theorem, we have also,

$$\begin{aligned}\mathbb{E}_{\mathbb{P}_\theta^i}(h(X^i)) &= \int_{\mathbb{R}} g(\varphi,\theta)\mathbb{E}_{\mathbb{P}^i}\left(h(X^i)L_{T_i}(X^i,\varphi,\varphi_0)\right)d\nu(\varphi)\\ &= \mathbb{E}_{\mathbb{P}^i}\left[h(X^i)\int_{\mathbb{R}} g(\varphi,\theta)L_{T_i}(X^i,\varphi,\varphi_0)d\nu(\varphi)\right].\end{aligned}$$

Hence, $\dfrac{d\mathbb{P}_\theta^i}{d\mathbb{P}^i} = \displaystyle\int L_{T_i}(X^i,\varphi,\varphi_0)g(\varphi,\theta)d\nu(\varphi)$, $\mathbb{P}^i$-a.s. and the independence of the individuals X^i yields the likelihood function. □

Proof. **(Proof of Proposition 4.1.2)**
Let $\beta_s(X^i,\varphi,\varphi_0) = H_2^+(X^i(s))\left(H_1(X^i(s),\varphi) - H_1(X^i(s),\varphi_0)\right)$. Clearly,

$$L_{T_i}(X^i,\varphi,\varphi_0) = \exp\left(\int_0^{T_i}\beta_s(X^i,\varphi,\varphi_0)dW^i(s) - \frac{1}{2}\int_0^{T_i}\beta_s(X^i,\varphi,\varphi_0)^2 ds\right).$$

Let $k_\varphi = \left(L_{T_i}(X^i,\varphi,\varphi_0)|_{T_i=t}\ ,\ t\in[0,T_i]\right)$. For each $\varphi\in\mathbb{R}^p$, k_φ is a supermartingale permitting the following representation: $k_\varphi(t) = 1 + \int_0^t k_\varphi(s)\beta_s(X^i,\varphi,\varphi_0)dW^i(s)$ which follows immediately from the Itô formula. Applying Fubini theorem for stochastic integrals (see, [73, theorem 5.15]), we assert that $\Lambda_i(X^i,\theta) = 1 + \int_0^{T_i}\gamma_s(X^i,\theta)dW^i(s)$, with

$$\gamma_s(X^i,\theta) = \int L_s(X^i,\varphi,\varphi_0)\beta_s(X^i,\varphi,\varphi_0)g(\varphi,\theta)d\nu(\varphi). \tag{6.3}$$

The measurability of $X^i \longrightarrow \gamma_s(X^i,\varphi_0)$ on C_{T_i} follows from the joint measurability of the following two functions:

- The function $(X^i,\varphi) \to \beta_s(X^i,\varphi,\varphi_0)$ which is measurable since $H_1(x)$ and $H_2(x,\varphi)$ are non anticipative functionals.

- The function $(X^i,\varphi) \to L_s(X^i,\varphi,\varphi_0)$ is also measurable (see, [25, Proposition 2]).

Now, as X^i are strong solutions, then $\Lambda_i(\theta)$ are $\mathcal{F}^{W(T_i)}$-measurable. Applying [73, Theorem 5.9], we claim that there exists a process $\left(\Psi^i(t,\theta),\ \mathcal{F}^{W^i(t)}\right)_{0\le t\le T_i}$ such that $\mathbb{P}\left(\int_0^{T_i}\Psi^i(t,\theta)^2 dt < \infty\right) = 1$ and for all $t \le T_i$,

$$\begin{aligned}\mathbb{E}\left(\Lambda_i(X^i,\theta)|\mathcal{F}^{W^i(t)}\right) &= \exp\left[\int_0^t \Psi^i(s,\theta)dW^i(s) - \frac{1}{2}\int_0^t \Psi^i(s,\theta)^2 ds\right] \\ &\quad \times \mathbb{E}(\Lambda_i(X^i,\theta)),\end{aligned}$$

where

$$\Psi^i(t,\theta) = \frac{\gamma_t(X^i,\theta)}{1+\displaystyle\int_0^t \gamma_s(X^i,\theta)dW^i(s)},$$

where $\gamma_t(X^i,\theta)$ is given by (6.3). Under the condition $\mathbb{E}\left(\int_0^{T_i}\gamma_s^2(X^i,\varphi_0)ds\right) < \infty$ we can use the martingale property, thus $\Lambda_i(X^i,\theta) = \exp\left[\int_0^T \Psi^i(s,\theta)dW^i(s) - \frac{1}{2}\int_0^T \Psi^i(s,\theta)^2 ds\right]$ and the log-likelihood function would be of the form:

$$\Gamma(\theta) = \sum_{i=1}^N \int_0^{T_i} \Psi^i(s,\theta)dW^i(s) - \frac{1}{2}\int_0^{T_i} \Psi^i(s,\theta)^2 ds.$$

□

Proof. **(Proof of the statement (4.7))** Let $\theta \in \Theta$ and set $\mathcal{J}_X(\theta_0,\theta) = \mathbb{E}_{\theta_0}\log\dfrac{\Lambda_1^{(\varepsilon)}(X^1,\theta_0)}{\Lambda_1^{(\varepsilon)}(X^1,\theta)}$. We simplify notations by setting $\Lambda_1(\theta) = \Lambda_1(X^1,\theta)$. $\mathcal{J}_X(\theta_0,\theta)$ vanishes if and only if $\theta = \theta_0$ (this can be easily justified by the identifiability assumption). We wish to show that $\mathcal{J}_X(\theta_0,\theta)$ is nonnegative for $\theta \ne \theta_0$. Set $A_{\theta,\theta_0} = \{\Lambda_1(\theta_0) \ge \Lambda_1(\theta)\}$. For each θ we have

$$\begin{aligned}\mathcal{J}_X(\theta_0,\theta) &= \mathbb{E}_{\theta_0}\log\frac{\Lambda_1^{(\varepsilon)}(X^1,\theta_0)}{\Lambda_1^{(\varepsilon)}(X^1,\theta)} \\ &= \mathbb{E}_{\theta_0}\log\frac{\varepsilon+\Lambda_1(\theta_0)}{\varepsilon+\Lambda_1(\theta)} \\ &\ge \mathbb{E}_{\theta_0}\left\{\log\left[\frac{\varepsilon+\Lambda_1(\theta_0)}{\varepsilon+\Lambda_1(\theta)}\right]\chi_{A_{\theta,\theta_0}}\right\} + \mathbb{E}_{\theta_0}\left\{\log\left[\frac{\varepsilon+\Lambda_1(\theta_0)}{\varepsilon+\Lambda_1(\theta)}\right]\chi_{A^c_{\theta,\theta_0}}\right\} \\ &\ge \mathbb{E}_{\theta_0}\left\{\log\left[\frac{\varepsilon+\Lambda_1(\theta_0)}{\varepsilon+\Lambda_1(\theta)}\right]\chi_{A^c_{\theta,\theta_0}}\right\}.\end{aligned}$$

We shall show that the RHS of the last inequality is nonnegative. Using the fact that, $\log\left(\dfrac{x+a}{x+b}\right)\chi_{(a<b)} > \log\dfrac{a}{b}\chi_{(a<b)}$, for all $a, b \in \mathbb{R}_+$, and $x \in (-a, \infty)$, we have

$$\mathbb{E}_{\theta_0}\left\{\log\left[\frac{\varepsilon+\Lambda_1(\theta_0)}{\varepsilon+\Lambda_1(\theta)}\right]\chi_{A^c_{\theta,\theta_0}}\right\} \geq \mathbb{E}_{\theta_0}\left\{\log\left[\frac{\Lambda_1(\theta_0)}{\Lambda_1(\theta)}\right]\chi_{A^c_{\theta,\theta_0}}\right\}. \tag{6.4}$$

If $\mathbb{P}_{\theta_0}(A^c_{\theta,\theta_0}) = 1$, the RHS of (6.4) is exactly the Kullback-Leibler information. Thereby $\mathcal{J}_X(\theta_0,\theta) \geq \mathcal{I}_X(\theta_0,\theta) > 0$. If $\mathbb{P}_{\theta_0}(A_{\theta,\theta_0}) = 1$, we have also $\mathcal{J}_X(\theta_0,\theta) > 0$. In general, using the fact that, $-\log(x) \geq 2 - 2\sqrt{x}$, for all $x > 0$, we develop the RHS of (6.4) as

$$\begin{aligned}
\mathbb{E}_{\theta_0}\left\{\log\left[\frac{\Lambda_1(\theta_0)}{\Lambda_1(\theta)}\right]\chi_{A^c_{\theta,\theta_0}}\right\} &= -\int_{A^c_{\theta,\theta_0}}\log\frac{\Lambda_1(\theta)}{\Lambda_1(\theta_0)}(\omega)d\mathbb{P}^1_{\theta_0}(\omega)\\
&\geq 2 - 2\int_{A^c_{\theta,\theta_0}}\sqrt{\frac{\Lambda_1(\theta)}{\Lambda_1(\theta_0)}}(\omega)\Lambda_1(\theta_0)(\omega)d\mathbb{P}^1(\omega)\\
&\geq 2 - 2\int_{A^c_{\theta,\theta_0}}\sqrt{\Lambda_1(\theta)\Lambda_1(\theta_0)}(\omega)d\mathbb{P}^1(\omega)\\
&\geq \mathbb{P}^1_{\theta}(A_{\theta,\theta_0}) + \mathbb{P}^1_{\theta_0}(A_{\theta,\theta_0})\\
&\qquad + \int_{A^c_{\theta,\theta_0}}\left[\sqrt{\Lambda_1(\theta)}(\omega) - \sqrt{\Lambda_1(\theta_0)}(\omega)\right]^2 d\mathbb{P}^1(\omega).
\end{aligned}$$

Hence, the proof of (4.7) is complete. □

Proof. **(Proof of the statement (4.9))** First, we recall some notations which will be used. $\mathbb{P}^1_{\theta_0}$ denotes the marginal distribution of the process X^1 on $(C_{T_1}, \mathcal{B}_{T_1})$ dominated by $\mathbb{P}^1 = \mu_{X^1_{\varphi_0,X^1}}$ with

$$\begin{aligned}
\frac{d\mathbb{P}^1_{\theta_0}}{d\mathbb{P}^1}(X^1) &= \int L_T(X^1,\varphi,\varphi_0)g(\varphi,\theta_0)d\nu(\varphi),\\
\text{and}\quad L_T(X^1,\varphi,\varphi_0) &= \frac{d\mu_{X^1_{\varphi,X^1}}}{d\mu_{X^1_{\varphi_0,X^1}}}(X^1)\quad \text{(see Subsection 4.1.2)}.
\end{aligned}$$

For any functional $h(X^1)$, we have

$$\begin{aligned}
\mathbb{E}_{\theta_0}(h(X^1)) &= \int_{C_{T_1}} h(x)\Lambda_1(x,\theta_0)d\mu_{X^1_{\varphi_0,x^1}}(x)\\
&= \int\int_{C_{T_1}} h(x)d\mu_{X^1_{\varphi,x^1}}(x)g(\varphi,\theta_0)d\nu(\varphi).
\end{aligned}$$

On the other hand, the expectation of a functional $H(X^1,\phi_1)$ under $\mathbb{P}$, with $\phi_1 \sim g(\varphi,\theta_0)d\nu(\varphi)$ is given by

$$\begin{aligned}
\mathbb{E}H(X^1,\phi_1) &= \int H(x,\varphi)d\mu_{(X^1,\phi_1)}(x,\varphi)\\
&= \int\int_{C_{T_1}} H(x,\varphi)d\mu_{X^1_{\varphi,x^1}}(x)g(\varphi,\theta_0)d\nu(\varphi)
\end{aligned}$$

Set $A_s(X^1) = \left|X^1(s)\right|^{2k}$. We have

$$\begin{aligned}\mathbb{E}_{\theta_0}\int_0^{T_1} A_s(X^1)ds &\leq T_1 \sup_{s\leq T_1}\mathbb{E}_{\theta_0}A_s(X^1)\\ &\leq T_1\sup_{s\leq T_1}\int\int_{C_{T_1}} A_s(x)d\mu_{X^1_{\varphi,x^1}}(x)g(\varphi,\theta_0)d\nu(\varphi)\\ &\leq T_1\sup_{s\leq T_1}\mathbb{E}A_s(X^1) = T_1\sup_{s\leq T_1}\mathbb{E}\left|X^1(s)\right|^{2k} < \infty. \end{aligned} \tag{6.5}$$

In the last inequality $A_s(X^1)$ is considered as functional of X^1 and ϕ_1; and the fact that $\mathbb{E}\left|\phi_1\right|^{2k} < \infty$ which in turn implies $\sup_{s\leq T_1}\mathbb{E}\left|X^1(s)\right|^{2k} < \infty$ (see Theorem 2.1.16 and note that (x^1,ϕ_1) is the initial condition of (4.1) with $i=1$).

□

Proof. **(Proof of Proposition 4.1.5)** In what follows, we will systematically use BDG inequality. We simplify notations by setting $\Lambda_1(\theta) = \Lambda^{(\varepsilon)}(X^1,\theta)$, $W_t = W^1(t)$, $\psi_k(t,\theta) = \dfrac{\partial\gamma_t(X^1,\theta)}{\partial\theta_k}$, $\psi_{kj}(t,\theta) = \dfrac{\partial^2\gamma_t(X^1,\theta)}{\partial\theta_k\partial\theta_j}$, and $\psi_{lkj}(t,\theta) = \dfrac{\partial^3\gamma_t(X^1,\theta)}{\partial\theta_l\partial\theta_k\partial\theta_j}$. **(i)** First, note that

$$\left|\frac{\partial}{\partial\theta_k}\log\Lambda_1(\theta)\right| \leq \varepsilon^{-1}\left|\int_0^T\psi_k(t,\theta)dW_t\right|,$$

which implies

$$\begin{aligned}\mathbb{E}_{\theta_0}\left\{\frac{\partial\log\Lambda_1(\theta)}{\partial\theta_k}\frac{\partial\log\Lambda_1(\theta)}{\partial\theta_j}\right\}^2 &\leq \frac{1}{2}\left\{\mathbb{E}_{\theta_0}\left(\frac{\partial\log\Lambda_1(\theta)}{\partial\theta_k}\right)^4 + \mathbb{E}_{\theta_0}\left(\frac{\partial\log\Lambda_1(\theta)}{\partial\theta_j}\right)^4\right\}\\ &\leq \frac{C_4}{2\varepsilon^4}\left\{\mathbb{E}_{\theta_0}\left(\int_0^T\psi_k(t,\theta)^2dt\right)^2 + \mathbb{E}_{\theta_0}\left(\int_0^T\psi_j(t,\theta)^2dt\right)^2\right\} < \infty.\end{aligned}$$

In particular $\mathcal{I}_X^{(\varepsilon)}(\theta_0)$ is finite. **(ii)** Let $\theta\in\overline{B_r(\theta_0)}$. For sufficiently small r, the mean value theorem yields for each $k,j\in\{1,\cdots,d\}$

$$\sup_{\theta\in\overline{B_r(\theta_0)}}\left|\frac{\partial^2\log\Lambda_1(\theta_0)}{\partial\theta_k\partial\theta_j} - \frac{\partial^2\log\Lambda_1(\theta)}{\partial\theta_k\partial\theta_j}\right| \leq r\sum_{l=1}^d\left|\frac{\partial^3\log\Lambda_1(\theta)}{\partial\theta_l\partial\theta_k\partial\theta_j}|_{\theta=\theta^*}\right|,$$

where θ^* is the maximizer of $\displaystyle\sum_{l=1}^d\left|\frac{\partial^3\log\Lambda_1(\theta)}{\partial\theta_l\partial\theta_k\partial\theta_j}\right|$ which depends only on $\overline{B_r(\theta_0)}$ (that is, it depends only on θ_0). Set $H(X^1,\theta_0) = \displaystyle\sum_{l=1}^d\left|\frac{\partial^3\log\Lambda_1(\theta)}{\partial\theta_l\partial\theta_k\partial\theta_j}|_{\theta=\theta^*}\right|$. We will prove that

$\mathbb{E}_{\theta_0} H(X^1, \theta_0) < \infty$. First, note that

$$
\begin{aligned}
\frac{\partial^3 \log \Lambda_1(\theta)}{\partial\theta_l \partial\theta_k \partial\theta_j} &= \sum_{i=1}^{5} G_i(\theta), \\
\text{where} \quad G_1(\theta) &= \Lambda_1(\theta)^{-1} \frac{\partial^3 \Lambda_1(\theta)}{\partial\theta_l \partial\theta_k \partial\theta_j} \\
G_2(\theta) &= -\Lambda_1(\theta)^{-2} \frac{\partial^2 \Lambda_1(\theta)}{\partial\theta_l \partial\theta_k} \frac{\partial \Lambda_1(\theta)}{\partial\theta_j} \\
G_3(\theta) &= -\Lambda_1(\theta)^{-2} \frac{\partial^2 \Lambda_1(\theta)}{\partial\theta_l \partial\theta_j} \frac{\partial \Lambda_1(\theta)}{\partial\theta_k} \\
G_4(\theta) &= -\Lambda_1(\theta)^{-2} \frac{\partial^2 \Lambda_1(\theta)}{\partial\theta_k \partial\theta_j} \frac{\partial \Lambda_1(\theta)}{\partial\theta_l} \\
G_5(\theta) &= 2\Lambda_1(\theta)^{-3} \frac{\partial \Lambda_1(\theta)}{\partial\theta_l} \frac{\partial \Lambda_1(\theta)}{\partial\theta_k} \frac{\partial \Lambda_1(\theta)}{\partial\theta_j}.
\end{aligned}
$$

We are going to show that $\mathbb{E}_{\theta_0} |G_1(\theta)|$, $\mathbb{E}_{\theta_0} |G_2(\theta)|$ and $\mathbb{E}_{\theta_0} |G_5(\theta)|$ are finite for all $\theta \in \overline{B_r(\theta_0)}$ and $0 < r < r_0$.

$$
\begin{aligned}
\mathbb{E}_{\theta_0} |G_1(\theta)| &\le \varepsilon^{-1} \mathbb{E}_{\theta_0} \left| \int_0^T \psi_{lkj}(t,\theta) dW_t \right| \\
&\le \varepsilon^{-1} \left[\mathbb{E}_{\theta_0} \left(\int_0^T \psi_{lkj}(t,\theta) dW_t \right)^2 \right]^{1/2} \\
&\le \frac{\sqrt{C_2}}{\varepsilon} \left[\mathbb{E}_{\theta_0} \int_0^T \psi_{lkj}(t,\theta)^2 dt \right]^{1/2} < \infty,
\end{aligned}
$$

where C_2 is a nonnegative due to BDG inequality. Hence $\mathbb{E}_{\theta_0} |G_1(\theta)|$ is finite.

$$
\begin{aligned}
\mathbb{E}_{\theta_0} |G_2(\theta)| &\le \varepsilon^{-2} \mathbb{E}_{\theta_0} \left| \int_0^T \psi_{lk}(t,\theta) dW_t \int_0^T \psi_j(t,\theta) dW_t \right| \\
&\le \frac{1}{2\varepsilon^2} \left\{ \mathbb{E}_{\theta_0} \left(\int_0^T \psi_{lk}(t,\theta) dW_t \right)^2 + \mathbb{E}_{\theta_0} \left(\int_0^T \psi_j(t,\theta) dW_t \right)^2 \right\} \\
&\le \frac{C_2}{2\varepsilon^2} \left\{ \mathbb{E}_{\theta_0} \int_0^T \psi_{lk}(t,\theta)^2 dt + \mathbb{E}_{\theta_0} \int_0^T \psi_j(t,\theta)^2 dt \right\} < \infty.
\end{aligned}
$$

Finally, we have

$$\mathbb{E}_{\theta_0}\left|G_5(\theta)\right| \le 2\varepsilon^{-3}\mathbb{E}_{\theta_0}\left|\int_0^T \psi_l(t,\theta)dW_t\int_0^T \psi_k(t,\theta)dW_t\int_0^T \psi_j(t,\theta)dW_t\right|$$
$$\le \varepsilon^{-3}\left\{\mathbb{E}_{\theta_0}\left(\int_0^T \psi_l(t,\theta)dW_t\right)^2 + \mathbb{E}_{\theta_0}\left(\int_0^T \psi_k(t,\theta)dW_t\int_0^T \psi_j(t,\theta)dW_t\right)^2\right\}$$
$$\le \varepsilon^{-3}\left\{\mathbb{E}_{\theta_0}\int_0^T \psi_l(t,\theta)^2dt + \frac{C_4}{2}\mathbb{E}_{\theta_0}\left(\int_0^T \psi_k(t,\theta)^2dt\right)^2 + \frac{C_4}{2}\mathbb{E}_{\theta_0}\left(\int_0^T \psi_j(t,\theta)^2dt\right)^2\right\} < \infty,$$

where C_4 is also a nonnegative constant due to BDG inequality. □

Proof. **(Proof of Proposition 4.1.7)** In this proof, we will systematically use BDG and Jensen inequalities, and Fubini theorem. **(i)** Set $\Lambda_i^{(\varepsilon)}(\theta) = \Lambda^{(\varepsilon)}(X^i,\theta)$. Since $\left|\beta(X^i(s),\varphi,\varphi_0)\right| \le \beta_s(\varphi)$, where $\beta_s(\varphi)$ is nonnegative deterministic function, we have

$$\mathbb{E}_{\theta_0}\left(L_s(X^i,\varphi,\varphi_0)\beta(X^i(s),\varphi,\varphi_0)\right)^2 \le$$
$$\le \beta_s(\varphi)^2\mathbb{E}_{\theta_0}\exp\left(2\int_0^s \beta(X^i(u),\varphi,\varphi_0)dW^i(u) - \int_0^s \beta(X^i(u),\varphi,\varphi_0)^2du\right)$$
$$\le \beta_s(\varphi)^2\left[\mathbb{E}_{\theta_0}\exp\left(4\int_0^s \beta(X^i(u),\varphi,\varphi_0)dW^i(u) - \frac{16}{2}\int_0^s \beta(X^i(u),\varphi,\varphi_0)^2du\right)\right]^{1/2} \times\left[\mathbb{E}_{\theta_0}\exp\left(6\int_0^s \beta(X^i(u),\varphi,\varphi_0)^2du\right)\right]^{1/2}$$
$$\le \beta_s(\varphi)^2\exp\left(3\int_0^T \beta_u(\varphi)^2du\right).$$

Thus

$$\begin{aligned}\mathbb{E}_{\theta_0}\int_0^T \gamma_s(X^i,\theta)^2ds &\le \int_0^T\int \mathbb{E}_{\theta_0}\left(L_s(X^i,\varphi,\varphi_0)\beta(X^i(s),\varphi,\varphi_0)\right)^2 g(\varphi,\theta)d\nu(\varphi)ds\\ &\le \int\int_0^T \beta_s(\varphi)^2ds\ \exp\left(3\int_0^T \beta_u(\varphi)^2du\right)g(\varphi,\theta)d\nu(\varphi)\\ &\le \int e^{4\int_0^T \beta_u(\varphi)^2du}g(\varphi,\theta)d\nu(\varphi) < \infty,\end{aligned}$$

where we used Cauchy Schwarz ineguality and the martingale property in the last inequality. For the continuity of the function $\Lambda_i^{(\varepsilon)}(\theta)$, we use the Kolmogorov-Chentsov criterion (2.3) with constants to be specified later. Let $\theta_1,\theta_2 \in N_\theta$, where N_θ is an open set such that

(4.13) holds true. We have

$$\begin{aligned}
&\mathbb{E}_{\theta_0}\left|\Lambda_i^{(\varepsilon)}(\theta_1)-\Lambda_i^{(\varepsilon)}(\theta_2)\right|^{2r} = (1+\varepsilon)^{-2r}\mathbb{E}_{\theta_0}\left|\Lambda(X^i,\theta_1)-\Lambda(X^i,\theta_2)\right|^{2r} \\
&= (1+\varepsilon)^{-2r}\mathbb{E}_{\theta_0}\left(\int L_T(X^i,\varphi,\varphi_0)\left[g(\varphi,\theta_1)-g(\varphi,\theta_2)\right]d\nu(\varphi)\right)^{2r} \\
&\leq \frac{C^{2r}}{(1+\varepsilon)^{2r}}\left\|\theta_1-\theta_2\right\|^{2\delta r}\int \mathbb{E}_{\theta_0}(L_T(X^i,\varphi,\varphi_0))^{2r}h(\varphi)d\nu(\varphi) \\
&\leq \frac{C^{2r}}{(1+\varepsilon)^{2r}}\left\|\theta_1-\theta_2\right\|^{2\delta r}\int e^{\lambda\int_0^T\beta_u(\varphi)^2du}h(\varphi)d\nu(\varphi)<\infty, \text{with } \lambda = 4r^2-r.
\end{aligned}$$

Once again, we used Cauchy Schwarz ineguality and the martingale property in the last inequality. By choosing r so that $\varepsilon = 2\delta r - d > 0$, we conclude that $\Lambda_i^{(\varepsilon)}(\theta)$ is continuous.

APPENDIX B.

Let $\mathcal{G}_n^+$ denote the class of nonnegative random sequences $\{X_i\}_{i=1}^n$ on the probability space $(\Omega, \mathcal{F}, \mathbb{P})$ such that for all $m > 0$, $\mathbb{E}X_i^m < \infty$. Let $\mathcal{P}\left(\mathcal{G}_n^+\right)$ be the subclass of $\mathcal{G}_n^+$ for which the following statement holds:

$$\mathbb{E}\prod_{i=1}^n X_i \leq \left(\sup_{1\leq i\leq n} \mathbb{E}X_i^{2^n}\right)^{n/2^n}.$$

Lemma 6.0.8. **(i)** *For all $m > 0$ and $n \in \mathbb{N}^*$, ${\mathcal{G}_n^+}^m \subset \mathcal{G}_n^+$, where ${\mathcal{G}_n^+}^m := \{u^m \,;\, u \in \mathcal{G}_n^+\}$.*

(ii) *For all $n \in \mathbb{N}^*$, $\mathcal{G}_n^+ = \mathcal{P}\left(\mathcal{G}_n^+\right)$.*

Proof. The first statement **(i)** is trivial. **(ii)** By induction on n, we shall show that $\mathcal{G}_n^+ \subset \mathcal{P}\left(\mathcal{G}_n^+\right)$ for all $n \geq 1$. $\mathcal{G}_1^+ \subset \mathcal{P}\left(\mathcal{G}_1^+\right)$ follows immediately from Cauchy Schwarz inequality. Assume that $\mathcal{G}_n^+ \subset \mathcal{P}\left(\mathcal{G}_n^+\right)$ and let $\{Y_i\}_{i=1}^{n+1} \in \mathcal{G}_{n+1}^+$. By using Cauchy Schwarz and Lyapunov inequalities, respectively, we obtain

$$\begin{aligned}
\mathbb{E}\prod_{i=1}^{n+1} Y_i &= \mathbb{E}Y_1 \prod_{i=1}^{n} Y_{i+1} \\
&\leq \left(\mathbb{E}Y_1^2\right)^{1/2} \left(\mathbb{E}\prod_{i=1}^{n} Y'_{i+1}\right)^{1/2} \\
&\leq \left(\mathbb{E}Y_1^{2^{n+1}}\right)^{1/2^{n+1}} \left(\mathbb{E}\prod_{i=1}^{n} Y'_{i+1}\right)^{1/2},
\end{aligned}$$

where $Y'_{i+1} = Y_{i+1}^2,\ i = 1, \cdots, n$. By **(i)** and hypothesis we have $\left\{Y'_{i+1}\right\}_{i=1}^n \in {\mathcal{G}_n^+}^2 \subset \mathcal{G}_n^+ \subset \mathcal{P}\left(\mathcal{G}_n^+\right)$, which yields

$$\begin{aligned}
\mathbb{E}\prod_{i=1}^{n+1} Y_i &\leq \left(\mathbb{E}Y_1^{2^{n+1}}\right)^{1/2^{n+1}} \left(\sup_{1\leq i\leq n} \mathbb{E}{Y'_{i+1}}^{2^n}\right)^{n/2^{n+1}} \\
&\leq \left(\mathbb{E}Y_1^{2^{n+1}}\right)^{1/2^{n+1}} \left(\sup_{2\leq i\leq n+1} \mathbb{E}Y_i^{2^{n+1}}\right)^{n/2^{n+1}} \\
&\leq \left(\sup_{1\leq i\leq n+1} \mathbb{E}Y_i^{2^{n+1}}\right)^{\frac{n+1}{2^{n+1}}}.
\end{aligned}$$

Hence $\mathcal{G}_{n+1}^+ \subset \mathcal{P}\left(\mathcal{G}_{n+1}^+\right)$ and the proof is complete. □

Proposition 6.0.9. *Under the weakned assumptions **A'**$_4$-**A'**$_5$, the individual density $\Lambda_i(X^i, \theta)$ given in (4.15) is continuous in θ $\mathbb{P}_{\theta_0}$-a.s.*

Proof. Once again, we simplify notations by omitting indices. Let $\theta_1,\ \theta_2 \in N_\theta$ (N_θ is an open set containing θ). By using Lemma 6.0.8 and the fact that $|a+b+c|^r \leq 2^{2r-2}(|a|^r + |b|^r + |c|^r)$, for all a, b, $c \in \mathbb{R}$ and $r \geq 1$, we obtain

$$\begin{aligned}
\mathbb{E}_{\theta_0} \prod_{l=1}^{n} \beta(X(t_l), \varphi, \varphi_0)^{2r} &\leq M^{2nr} \mathbb{E}_{\theta_0} \prod_{l=1}^{n} (1 + |X(t_l)|^{\gamma_1} + \|\varphi\|^{\gamma_2})^{2r} \\
&\leq M^{2nr} \left[\sup_{t \leq T} \mathbb{E}_{\theta_0} (1 + |X_t|^{\gamma_1} + \|\varphi\|^{\gamma_2})^{2^{n+1}r} \right]^{n/2^n} \\
&\leq M^{2nr} \left[2^{2(2^{n+1}r)-2} \right]^{n/2^n} \left[\sup_{t \leq T} \mathbb{E}_{\theta_0} \left(1 + |X_t|^{\gamma_1 2^{n+1} r} + \|\varphi\|^{\gamma_2 2^{n+1} r} \right) \right]^{n/2^n} \\
&\leq M^{2nr} 2^{4nr - n/2^{n-1}} \left[1 + \sup_{t \leq T} \mathbb{E}_{\theta_0} |X_t|^{\gamma_1 nr} + \|\varphi\|^{2\gamma_2 nr} \right] \\
&\leq M^{2nr} 2^{4nr - n/2^{n-1}} \left[1 + M_3^{\gamma_1 nr} + \|\varphi\|^{2\gamma_2 nr} \right].
\end{aligned}$$

Thus

$$\begin{aligned}
\left\{ \mathbb{E} \left| \Lambda^m(X, \theta_1) - \Lambda^m(X, \theta_2) \right|^{2r} \right\}^{1/2r} &\leq C \|\theta_1 - \theta_2\|^\delta \sum_{n \geq 1} \frac{c^n}{\sqrt[2r]{n!}} \left\{ 1 + M_3^{\gamma_1 nr} + \int \|\varphi\|^{2\gamma_2 nr} h(\varphi) d\nu(\varphi) \right\}^{1/2r} \\
&\leq C \|\theta_1 - \theta_2\|^\delta \sum_{n \geq 1} \frac{c^n}{\sqrt[2r]{n!}} \left\{ 1 + M_3^{\gamma_1 n/2} + \mu_{n,\gamma_2,r}^{1/2r} \right\},
\end{aligned}$$

where $c = 4M \sqrt[2r]{T^r C_r}$. □

APPENDIX C.

Lemma 6.0.10. *For all $c > 0$ and $\alpha \in (0, 1)$, we have*

$$\mathbb{P}\left(|Z_1 + Z_2| > c\right) \leq \mathbb{P}\left(|Z_1| > (1-\alpha)c\right) + \mathbb{P}\left(|Z_2| > \alpha c\right),$$

where Z_1 and Z_2 are two random variables.

Lemma 6.0.11. *Let $(X_n,\ n \geq 0)$ be a random sequence that converges weakly to a random variable X. Let A be a Borel set such that $\mathbb{P}(X \in A) > 0$ and $\mathbb{P}(X \in \delta A) = 0$, where δA denotes the boundary of the set A. For sufficiently large n, we have*

$$\mathbb{E}\left|\chi_{(X_n \in A)} - \chi_{(X \in A)}\right| \leq \sqrt{2}\mathbb{P}(X \in A)^{1/2}\left\{\mathbb{P}(X_n \notin A)^{1/2} + \mathbb{P}(X \notin A)^{1/2}\right\}.$$

Proof. Simple computations yield

$$\begin{aligned}
\mathbb{E}\left|\chi_{(X_n \in A)} - \chi_{(X \in A)}\right| &= \mathbb{E}\left(\chi_{(X_n \in A)} - \chi_{(X \in A)}\right)^2 \\
&= \mathbb{P}\left(X \in A\right) - \mathbb{P}\left(X, X_n \in A\right) + \mathbb{P}\left(X_n \in A\right) - \mathbb{P}\left(X, X_n \in A\right) \\
&= \mathbb{E}\left\{\chi_{(X \in A)}\left(1 - \chi_{(X_n \in A)}\right)\right\} + \mathbb{E}\left\{\chi_{(X_n \in A)}\left(1 - \chi_{(X \in A)}\right)\right\} \\
&\leq \left[\mathbb{P}\left(X \in A\right)\mathbb{P}\left(X_n \notin A\right)\right]^{1/2} + \left[\mathbb{P}\left(X_n \in A\right)\mathbb{P}\left(X \notin A\right)\right]^{1/2}. \qquad (6.6)
\end{aligned}$$

(6.6) is justified by the Cauchy-Schwarz inequality. Since $\{X_n\}_{n \geq 0}$ converges weakly to X, then by the Portmanteau lemma (e.g., [113]) we have

$$\mathbb{P}\left(X_n \in A\right) \leq 2\mathbb{P}(X \in A), \qquad (6.7)$$

for all $n \geq n_0$, where n_0 is sufficiently large. The desired result follows from (6.6) and (6.7). □

Lemma 6.0.12. *Let $X_i,\ i = 1, \cdots, N$, be a sequence of i.i.d random variables with common density f. Assume that f is continuous with compact support $A \subset \mathbb{R}$. Let $A_j(h) = [hj, h(j+1))$, $j = 1, \cdots, J$ denote all Borel sets for which $\lambda\left(A_j(h) \cap A\right) \neq 0$. We have*

$$\lim_{h \to 0} \frac{1}{N} \sum_{i=1}^{N} \sum_{j=1}^{J} \mathbb{P}\left(X_i \in A_j(h)\right)^{1/2} = 0.$$

Proof. Actually,

$$\begin{aligned}
\sup_{i,j} \mathbb{P}\left(X_i \in A_j(h)\right) &= \sup_{j} \int_{hj}^{h(j+1)} f(t)dt \\
&\leq \sup_{t} f(t)h \to 0 \text{ as } h \to 0.
\end{aligned}$$

Let $\varepsilon > 0$. There exists $h_0 > 0$ such that $\mathbb{P}\left(X_i \in A_j(h)\right) < \varepsilon^2/j^4$, for all i and $h \in (0, h_0)$. Hence

$$\frac{1}{N} \sum_{i=1}^{N} \sum_{j=1}^{J} \mathbb{P}\left(X_i \in A_j(h)\right)^{1/2} \leq \varepsilon \left(\sum_{j \geq 1} \frac{1}{j^2}\right) < \infty.$$

□

APPENDIX D.

Lemma 6.0.13. *Let the assumptions $\boldsymbol{A}_1$ and $\boldsymbol{A}_2$ (or $\boldsymbol{A}$') be satisfied. We have*

(i) *For $i = 1, \cdots, N$ and $t \geq 0$, we have*

$$\left|X^i(t) - x^i(t)\right| \leq \varepsilon \left(\left|W^{i,H}(t)\right| + L(|\phi_i| \vee 1) \int_0^t e^{L(|\phi_i| \vee 1)(t-s)} \left|W^{i,H}(s)\right| ds \right). \tag{6.8}$$

(ii) *For $i = 1, \cdots, N$, we have*

$$\begin{aligned} \sup_{0 \leq t \leq T} \mathbb{E}\left(X^i(t) - x^i(t)\right)^2 &\leq 2\varepsilon^2 \left(T^{2H} + \frac{2e^{3LT}}{2H+1} \mu_f(3LT) T^{2H} \right) \\ &:= C(L, T, H, f)\varepsilon^2, \end{aligned}$$

where $\mu_f(\lambda)$ is defined by(5.21).

It is easy to see that (ii) in Lemma 6.0.13 can be generalized as follows

$$\sup_{0 \leq t \leq T} \mathbb{E}\left|X^i(t) - x^i(t)\right|^{2p} = O\varepsilon^{2p}, \quad \text{for all } i \text{ and } p \geq 1$$

Proof. We focus on one single subject X^1 and simplify notations by omitting indices. Let X_t and x_t be solutions to equations (5.20) and (5.22), respectively. For the statement (i), we fix $u_t = |X_t - x_t|$, so

$$\begin{aligned} u_t &= \left| \int_0^t (\phi b(X_s) - \phi b(x_s))\, ds + \varepsilon W_t^H \right| \\ &\leq \int_0^t |\phi b(X_s) - \phi b(x_s)|\, ds + \varepsilon \left|W_t^H\right| \\ &\leq L \int_0^t u_s ds + \varepsilon \left|W_t^H\right| \quad (\text{By } \mathbf{A}_1\ ). \end{aligned}$$

But, if the alternative condition **A'** is considered, we use the fact that $|b(x) - b(x')| \leq L|x - x'|$, for all $x, x' \in \mathbb{R}$ ($L := \sup_x |b'(x)|$). Hence,

$$u_t \leq L(|\phi| \vee 1) \int_0^t u_s ds + \varepsilon \left|W_t^H\right|.$$

By applying Gronwall inequality (see, [73, Lemma 4.15]) with $c = L(|\phi| \vee 1)$ and $v(t) = \varepsilon \left|W_t^H\right|$, the statement (6.8) follows imediately.

For the statement (ii), we set $\phi' = |\phi| \vee 1$ and apply (i) to get

$$\begin{aligned} \sup_{0 \leq t \leq T} \mathbb{E}\left(X_t - x_t\right)^2 &\leq 2\varepsilon^2 \sup_{0 \leq t \leq T} \left\{ \mathbb{E}(W_t^H)^2 + \mathbb{E}\left(L\phi' \int_0^t e^{L\phi'(t-s)} \left|W_s^H\right| ds \right)^2 \right\} \\ &\leq 2\varepsilon^2 \left\{ T^{2H} + \frac{2}{T} \mathbb{E}\left(\frac{(TL\phi')^2}{2} e^{2L\phi' T} \right) \int_0^T \mathbb{E}\left|W_s^H\right|^2 ds \right\} \\ &\leq 2\varepsilon^2 \left(T^{2H} + \frac{2e^{3LT}}{2H+1} \mu_f(3LT) T^{2H} \right), \end{aligned}$$

where $\mu_f(\lambda)$ is given in (5.21). In the last inequality, we used the fact that $\frac{x^2}{2} \leq e^x, \ \forall x \geq 0$. □

Lemma 6.0.14. *Let Q, Q_n, $n \geq 1$ be random variables on the same probability space $(\Omega, \mathcal{F}, P)$. Assume that $|Q_n - Q| = O_{\mathbb{P}}\delta_n$ with $\delta_n \longrightarrow 0$ as $n \to \infty$. Then,*

$$\left|\frac{Q_n}{Q} - 1\right| = O_{\mathbb{P}}\delta_n.$$

Proof. Let $c > 0$, we have

$$\begin{aligned}
\sup_n \mathbb{P}\left(\left|\frac{Q_n}{Q} - 1\right| > c\delta_n\right) &= \sup_n \mathbb{P}\left(\frac{1}{|Q|}\left[\delta_n^{-1}|Q_n - Q|\right] > c\right) \\
&\leq \sup_n \mathbb{P}\left(\frac{1}{2|Q|^2} + \frac{1}{2}\left[\delta_n^{-1}|Q_n - Q|\right]^2 > c\right) \\
&\leq \mathbb{P}\left(\frac{1}{2|Q|^2} > \frac{c}{2}\right) + \sup_n \mathbb{P}\left(\frac{1}{2}|Q_n - Q|^2 > \frac{c}{2}\delta_n^2\right) \\
&\leq \mathbb{P}\left(1 > c|Q|^2\right) + \sup_n \mathbb{P}\left(|Q_n - Q| > \sqrt{c}\,\delta_n\right) \\
&\longrightarrow 0 \quad \text{as } c \to \infty.
\end{aligned}$$

□

Lemma 6.0.15. ***(Tsybakov [110, p.14])*** *Assume that $f \in \mathcal{N}(\delta, R)$ and let G be a kernel of order $l = \lfloor \delta \rfloor$ satisfying*

$$\int_{\mathbb{R}} |u|^\delta |G(u)|\, du < \infty.$$

Then, for any $h_\varepsilon > 0$,

$$\|f_{h_\varepsilon} - f\|^2_{L^2(\mathbb{R})} \leq C(l, \delta, R) h_\varepsilon^{2\delta},$$

where $f_{h_\varepsilon}(x) = \frac{1}{h_\varepsilon}\int_{\mathbb{R}} G\left(\frac{x-u}{h_\varepsilon}\right)$ and $C(l, \delta, R) = \frac{R}{l!}\int_{\mathbb{R}} |u|^\delta |G(u)|\, du$.

Lemma 6.0.16. *Let G be a kernel which satisfies $\|G\|_{L^2(\mathbb{R})} < \infty$. We have*

$$\mathbb{E}\left\|\widehat{f}^{(1)}_{h_\varepsilon} - \mathbb{E}\widehat{f}^{(1)}_{h_\varepsilon}\right\|^2_{L^2(\mathbb{R})} \leq \frac{\|G\|^2_{L^2(\mathbb{R})}}{N h_\varepsilon}.$$

Proof. Set $\eta_{i,n}(x) = G_{h_\varepsilon}(x - \widehat{\phi}^{(1)}_{i,n}) - \mathbb{E}\left(G_{h_\varepsilon}(x - \widehat{\phi}^{(1)}_{i,n})\right)$, where $G_{h_\varepsilon}(u) = \frac{1}{h_\varepsilon}G\left(\frac{u}{h_\varepsilon}\right)$. $\eta_{i,n}(x)$, $i = 1, \cdots, N$ are i.i.d random variables with $\mathbb{E}\left[\eta_{1,n}(x)\right] = 0$, and with a change of variables $\frac{x - \widehat{\phi}^{(1)}_{1,n}}{h_\varepsilon} = y$ in the second inequality below, we get

$$\begin{aligned}\int_{\mathbb{R}} \mathbb{E}\left(\eta_{1,n}(x)\right)^2 dx &\le \int_{\mathbb{R}} \mathbb{E}\left(G_{h_\varepsilon}(x-\widehat{\phi}_{1,n}^{(1)})\right)^2 dx \\ &\le \frac{1}{h_\varepsilon^2}\mathbb{E}\int_{\mathbb{R}} G^2\left(\frac{x-\widehat{\phi}_{1,T}}{h_\varepsilon}\right) dx \\ &\le \frac{1}{h_\varepsilon}\int_{\mathbb{R}} G^2(y)dy.\end{aligned}$$

Therefore

$$\begin{aligned}\mathbb{E}\left(\left\|\widehat{f}_{h_\varepsilon}^{(1)}-\mathbb{E}(\widehat{f}_{h_\varepsilon}^{(1)})\right\|_{L^2(\mathbb{R})}^2\right) &= \mathbb{E}\int_{\mathbb{R}}\left(\widehat{f}_{h_\varepsilon}^{(1)}(x)-\mathbb{E}\widehat{f}_{h_\varepsilon}^{(1)}(x)\right)^2 dx \\ &= \frac{1}{N^2}\mathbb{E}\int_{\mathbb{R}}\left(\sum_{i=1}^{N}\eta_{i,n}(x)\right)^2 dx \\ &= \frac{1}{N}\int_{\mathbb{R}}\mathbb{E}\left(\eta_{1,n}(x)\right)^2 dx \le \frac{\|G\|_{L^2(\mathbb{R})}^2}{Nh_\varepsilon}.\end{aligned}$$

□

Lemma 6.0.17. *Under the assumptions **A'**$_3$ and **A'**$_4$, we have*

$$\sup_x \left|\widehat{f}_{h_\varepsilon}^{(1)}(x)-\widetilde{f}_{h_\varepsilon}(x)\right| = O_{\mathbb{P}}\delta_n h_\varepsilon^{-2}, \quad \text{as } \delta_n h_\varepsilon^{-2} \longrightarrow 0,$$

where $\delta_n = \varepsilon_n^{1/(3-2H)}$.

For the proof of this Lemma, we recall the following technical result used systematically througout this paper.

Lemma 6.0.18. *Lets $Z_1, \cdots, Z_N$ a sequence of random variables defined on common probability space $(\Omega, \mathcal{F}, \mathbb{P})$ and $\gamma_1, \cdots, \gamma_N$ a sequence of nonnegative real numbers such that with $\sum_{i=1}^{N}\gamma_i = 1$, then*

$$\mathbb{P}\left(\sum_{i=1}^{N}|Z_i| > c\right) \le \sum_{i=1}^{N}\mathbb{P}\left(|Z_i| > c\gamma_i\right).$$

Proof of Lemma 6.0.17 First, we recall that $G_{h_\varepsilon}(u) = \dfrac{1}{h_\varepsilon} G\left(\dfrac{u}{h_\varepsilon}\right)$. Let $(\lambda_n)_{n\geq 1}$ be a nonnegative sequence such that $\lambda_n \longrightarrow 0$; to be specified later. We have

$$\begin{aligned}
\left|\widehat{f}^{(1)}_{h_\varepsilon}(x) - \widetilde{f}_{h_\varepsilon}(x)\right| &\leq \frac{1}{N}\sum_{i=1}^{N}\left|G_{h_\varepsilon}(x-\widehat{\phi}^{(1)}_{i,n}) - G_{h_\varepsilon}(x-\phi_i)\right| \\
&\leq \frac{1}{Nh_\varepsilon^2}\sum_{i=1}^{N}|G'(x_i^*)|\left|\widehat{\phi}^{(1)}_{i,n} - \phi_i\right| \\
&\leq \frac{M'}{Nh_\varepsilon^2}\sum_{i=1}^{N}\left|\widehat{\phi}^{(1)}_{i,n} - \phi_i\right|,
\end{aligned}$$

where x_i^* is convex combination of $(x-\widehat{\phi}^{(1)}_{i,n})$ and $(x-\phi_i)$. In the second inequality above, we used the mean value theorem. Therefore, for any $c > 0$, the Lemma 6.0.18 with $\gamma_i = \dfrac{1}{N}$ gives

$$\begin{aligned}
\sup_n \mathbb{P}\left(\left|\widehat{f}^{(1)}_{h_\varepsilon}(x) - \widetilde{f}_{h_\varepsilon}(x)\right| > c\lambda_n\right) &\leq \sup_n \mathbb{P}\left(\frac{M'}{Nh_\varepsilon^2}\sum_{i=1}^{N}\left|\widehat{\phi}^{(1)}_{i,n} - \phi_i\right| > c\lambda_n\right) \\
&\leq \sum_{i=1}^{N}\mathbb{P}\left(\left|\widehat{\phi}^{(1)}_{i,n} - \phi_i\right| > c\lambda_n\left(\frac{Nh_\varepsilon^2}{M'}\right)\right) \\
&\leq \sum_{i=1}^{N}\mathbb{P}\left(\left|\widehat{\phi}^{(1)}_{i,n} - \phi_i\right| > c\delta_n\right)
\end{aligned}$$

The last inequality is justified by letting $\lambda_n = \dfrac{M'}{h_\varepsilon^2}\delta_n$ and $\delta_n = \varepsilon_n^{1/(3-2H)}$. Thus, using Theorem 5.4.3, we obtain

$$\sup_n \mathbb{P}\left(\left|\widehat{f}^{(1)}_{h_\varepsilon}(x) - \widetilde{f}_{h_\varepsilon}(x)\right| > c\lambda_n\right) \longrightarrow 0 \quad \text{as } c \longrightarrow \infty.$$

□

Lemma 6.0.19. *Under the assumptions **A'**$_3$ and **A'**$_4$, we have*

$$\sup_x \left|f(x) - \mathbb{E}\widetilde{f}_{h_\varepsilon}(x)\right| = O h_\varepsilon^\gamma.$$

Proof. First, note $\mathbb{E}[\widetilde{f}_{h_\varepsilon}(x)] = \mathbb{E}\left[G_{h_\varepsilon}(x-\phi_1)\right]$, since ϕ_i are i.i.d random variables. Hence, for any x

$$\begin{aligned}
\left|f(x) - \mathbb{E}[\widetilde{f}_{h_\varepsilon}(x)]\right| &= \left|\frac{1}{N}\sum_{i=1}^{N}\mathbb{E}[G_{h_\varepsilon}(x-\phi_i)] - f(x)\right| \\
&= \left|\mathbb{E}[G_{h_\varepsilon}(x-\phi_1)] - f(x)\int_{\mathbb{R}} G(v)dv\right| \\
&\leq \left|\int_{\mathbb{R}} G_{h_\varepsilon}(x-v)f(v)dv - \int_{\mathbb{R}} G(v)f(x)dv\right| \\
&\leq \left|\int_{\mathbb{R}} G(v)(f(x-vh_\varepsilon) - f(x))dv\right| \\
&\leq \int_{\mathbb{R}} |G(v)|\,|(f(x-vh_\varepsilon) - f(x))|\,dv \\
&\leq Dh_\varepsilon^\gamma \int_{\mathbb{R}} |v|^\gamma\,|G(v)|\,dv.
\end{aligned}$$

Since the last upper bound given above is independent of x, the proof is complete. □

Lemma 6.0.20. *Under **A'**$_4$, we have*

$$\sup_x \mathbb{E}\left|\widetilde{f}_{h_\varepsilon}(x) - \mathbb{E}\widetilde{f}_{h_\varepsilon}(x)\right|^2 = O\frac{h_\varepsilon^{-2}}{N}.$$

In particular,

$$\left|\widetilde{f}_{h_\varepsilon}(x) - \mathbb{E}\widetilde{f}_{h_\varepsilon}(x)\right| = O_{\mathbb{P}} h_\varepsilon^{-1}/\sqrt{N}, \quad \forall x \in \mathbb{R}.$$

Proof. Let $x \in \mathbb{R}$ and set $\eta_{h_\varepsilon,i}(x) = G_{h_\varepsilon}(x-\phi_i) - \mathbb{E}G_{h_\varepsilon}(x-\phi_i)$. Since $\eta_{h_\varepsilon,i}(x)$, $i = 1, \cdots, N$ are i.i.d random variables, we have

$$\begin{aligned}
\mathbb{E}\left|\widetilde{f}_{h_\varepsilon}(x) - \mathbb{E}\widetilde{f}_{h_\varepsilon}(x)\right|^2 &= \mathbb{E}\left(\frac{1}{N}\sum_{i=1}^{N}\eta_{h_\varepsilon,i}(x)\right)^2 \\
&= \frac{1}{N^2}\mathbb{E}(\eta_{h_\varepsilon,1}(x))^2 \\
&\leq \frac{1}{N}\mathbb{E}\left(G_{h_\varepsilon}(x-\phi_1)\right)^2 \\
&\leq \frac{1}{Nh_\varepsilon^2}\int_{\mathbb{R}} G^2\left(\frac{x-v}{h_\varepsilon}\right)f(v)dv \leq \frac{M^2}{Nh_\varepsilon^2}.
\end{aligned}$$

□